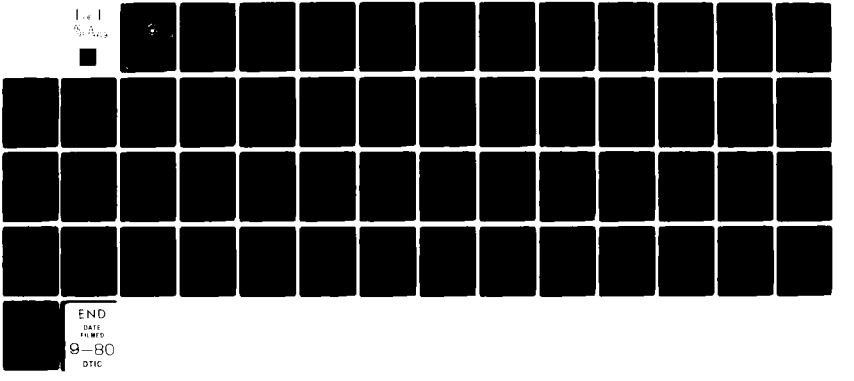
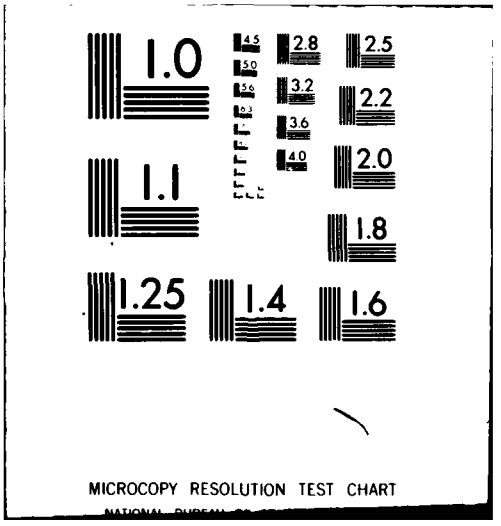
AD-A087 369 NAVAL POSTGRADUATE SCHOOL MONTEREY CA F/G 12/1
REGRESSION-ADJUSTED ESTIMATES FOR REGENERATIVE SIMULATIONS, WIT--ETC(U)
APR 80 P A LEWIS, P HEIDELBERGER
UNCLASSIFIED NPS55-80-019 NL



END
DATE FILMED
9-80
DTIC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
MICROCOPY RESOLUTION TEST CHART

NPS55-80-019

NAVAL POSTGRADUATE SCHOOL

Monterey, California

REGRESSION-ADJUSTED ESTIMATES FOR REGENERATIVE SIMULATIONS, WITH GRAPHICS

by

P. A. W. Lewis
and
P. Heidelberger

April 1980

Prepared for:

Naval Postgraduate School
Monterey, California 93940

NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA

Rear Admiral J. J. Ekelund
Superintendent

Jack R. Borsting
Provost

This report was prepared by:

P. Heidelberger by P.A.W.L

P. Heidelberger, Res. Assoc.
IBM Research Lab
Yorktown Heights, N.Y.

P.A.W. Lewis

P. A. W. Lewis, Professor
Department of Operations Research

Reviewed by:

G. Howard, Actg Ch

Michael G. Sovereign, Chairman
Department of Operations Research

Released by:

William M. Tolles

William M. Tolles
Dean of Research

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: NPS55-80-019	2. GOVT ACCESSION NO.: AD-A087369
	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): Regression-Adjusted Estimates for Regenerative Simulations, with Graphics.	5. TYPE OF REPORT & PERIOD COVERED: Techical rept.
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): P.A.W. Lewis and P. Heidelberger	8. CONTRACT OR GRANT NUMBER(s): 55
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Naval Postgraduate School, Monterey, Ca. 93940	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Naval Postgraduate School, Monterey, Ca. 93940	12. REPORT DATE: Apr 80
	13. NUMBER OF PAGES: 53
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The independent block structure of regenerative processes and the known convergence rates of the means of ratio estimators are exploited to produce bias-free regression-adjusted estimates (rare's) for regenerative simulations. Direct assessments of the variances of the estimates are obtained, as well as indications--both formal and graphical--of their normality or non-normality.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102-014-6601 |

UNCLASSIFIED.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

251450 gw

REGRESSION-ADJUSTED ESTIMATES FOR REGENERATIVE SIMULATIONS, WITH GRAPHICS

P. Heidelberger
IBM, T.J. Watson Research Center
Yorktown Heights, N.Y. 10598

P. A. W. Lewis*
Naval Postgraduate School
Monterey, Calif. 93940

ABSTRACT

The independent block structure of regenerative processes and the known convergence rates of the means of ratio estimators are exploited to produce bias-free regression-adjusted estimates (*rare*'s) for regenerative simulations. Direct assessments of the variances of the estimates are obtained, as well as indications--both formal and graphical--of normality and symmetry, or lack of it, in the distribution of the estimates.

1. INTRODUCTION

In regenerative simulations one exploits the fact that the sample path of the simulated process, say $\{W_i, i \geq 0\}$, can be divided into independent and identically distributed (i.i.d.) blocks of lengths $\{\tau_j, j \geq 1\}$. For example, let the quantity of interest be the expected stationary waiting time, E(W), in a queue and let Y_j be the sum of the waiting times observed in the jth block. Then the usual regenerative estimate, $re(n)$, of E(W), formed from n blocks is $re(n) = \bar{Y}/\bar{\tau}$ where $\bar{Y}$ and $\bar{\tau}$ are the averages of $\{Y_1,\ldots, Y_n\}$, and $\{\tau_1,\ldots, \tau_n\}$, respectively. This estimate converges to E(W) as n gets large and the so-called "statistics" of regenerative simulation exploit the fact that $re(n)$ is asymptotically unbiased and normally distributed with a variance of a known form which can be estimated from the data [Crane, and Iglehart (1975); Crane and Lemoine (1977); Iglehart (1978)].

An advantage of the regenerative estimate is that it eliminates the problem which arises in straightforward simulations of choosing initial conditions so as to make $\{W_i, i \geq 0\}$ stationary and $\bar{W}$, the sample path average, unbiased. It is crucial to understand that the initial conditions are of importance only for finite sample

*Work supported in part by the Office of Naval Research under Grant NR-42-84.

sizes so that the advantage of $\mathcal{re}(n)$ over $\bar{W}$ disappears asymptotically. Given this advantage, the regenerative estimate is usually employed and a single sample path is used for as long as one needs to simulate in order to achieve a given precision. At that point the simulator assumes that the point estimate $\mathcal{re}(n)$ is both unbiased and normally distributed. There are several problems involved in this "flying-blind" procedure besides the fact that the assumptions of normality and no bias are typically not verified and may not be true. Most notably, it is not known how to adjust confidence interval widths to allow for the fact that the variance is estimated from the data and is usually highly positively correlated with the point estimate. This correlation is particularly troublesome in sequential procedures which use a relative confidence-interval-width criterion for determining simulation run lengths.

A method which is preferable to using a single sample path would be to generate m i.i.d. sample paths, form an averaged estimate from the m estimates obtained and then produce the usual standard deviation estimate for a sample mean. Unfortunately taking m sample paths of length n/m only aggravates the concern with bias and normality of the estimate which occurs with both the straightforward and regenerative methods.

In this paper we exploit two aspects of the regenerative structure of the simulated process to resolve this dilemma. The first is that the regenerative blocks are i.i.d. so that a single simulation of n cycles can always be sectioned into k i.i.d. simulations of $n_k = n/k$ blocks per section for multiple values of k. Thus from a single simulation of n blocks, unbiased (but correlated) estimates of $E(\mathcal{re}(n_k))$ can be obtained for the multiple values of n_k. Secondly, unlike the estimate $\bar{W}$, the regenerative estimate has known bias structure, i.e., $E(\mathcal{re}(n)) = E(W) + \beta_1/n + \beta_2/n^2 + \cdots$. Combining these two facts and using regression techniques an estimate and a graphical picture of the bias structure can be obtained. More importantly, a <u>regression adjusted regenerative estimate</u>, $\mathcal{rare}(n)$, for $E(W)$) can be produced which corrects for the bias.

This procedure can itself be replicated, say, m times producing m i.i.d. estimates *rare*(1,n),...,*rare*(m,n) with low bias and the value E(W) can then be estimated by *arare*(m,n), the average of these m *rare*(j,n)'s. Furthermore plots showing the evolution to the unbiased state can be produced along with box plots displaying the distributions of the estimates. Running normal plots with normal test statistics can also be given for various subrun lengths n_k. These plots allow the user to judge the subsample length n at which *rare*(1,n),...,*rare*(m,n) can be assumed to be normally distributed. Using this sample of m *rare*'s a standard deviation estimate, *srare*(m,n), for the averaged *rares*, *arare*(m,n), is obtained from which a confidence interval for the steady state parameter is derived. Furthermore the sample of m, *rare*(1,n),..., *rare*(m,n), is available to do comparisons and rankings and to form multidimensional confidence regions (when more than one parameter is of interest) using standard normal theory statistical methodology or nonparametric methodology.

Other techniques which reduce bias in regenerative estimates are the jackknife and Tin estimates [Miller (1970); Iglehart (1975)] and estimates based on renewal theoretic properties [Heidelberger (1978), and Meketon (1979)]. As generally applied these techniques remove only the 1/n term in the bias expansion (the jackknife can be modified to remove the $1/n^2$ term at a great computational cost and an uncertain effect on the variance of the resulting estimate). However, no guidance is given as to when this bias reduction is sufficient and the jackknife is known to inflate the variance in finite samples [Efron and Stein (1978); Goodman, Lewis and Robbins (1971)]. Furthermore, the problems of correlation between the point and variance estimates and of detecting normality are not relieved by these alternative estimators. The regression and graphical procedures given in Sections 3 and 4 could be applied to estimates for which the 1/n term has already been removed, by, for instance, jackknifing, although this seems redundant. In practice the bias becomes insigificant long before the assumption of normality is valid, particularly when bias reducing techniques are applied.

The work in this paper is related to that of Fishman (1977) who used replications with the Tin estimate to reduce bias, and a normality test. The strength of the

regression adjusted regenerative estimate in conjunction with the graphics is that one obtains a very clear picture of how the bias is changing and at what point it is insignificant. In addition, the running normal plots show the convergence to normality.

To apply the methodology given in this paper in practice requires a relatively sophisticated user to interactively interpret the output. This sharply contrasts with previous automated sequential procedures [Fishman (1977) and Lavenberg and Sauer (1977)], which require no human intervention, but which provide the user with far less information, particularly about the dynamics of the simulation.

2. OBJECTIVES AND OUTLINE

The various regression adjusted estimates are described in Section 3 of the paper and the graphics which has been developed to go with them are described in Section 4. In Section 5 we describe a protocol which directs the simulator in the use of the above methodology.

The objective of this protocol is to produce, for a given precision or total available computing, estimates $\mathit{rare}(j,m)$, $j = 1,\ldots,m$, from the shortest possible number of blocks per replication such that the $\mathit{rare}(j,m)$'s are approximately

(i) unbiased;

(ii) normally distributed.

The minimization of the number of blocks in each replication is done in order to optimize the number, m, of replications of the simulation from which

(i) a standard deviation estimate $s\mathit{rare}(m,n)$ for the averaged rare estimates, $\mathit{arare}(m,n)$, is obtained;

(ii) a confidence interval for the parameter is obtained using $s\mathit{rare}(m,n)$ and $\mathit{arare}(m,n)$;

(iii) a sample of m $\mathit{rare}(j,n)$'s is obtained to do comparisons, rankings, etc. using standard normal theory statistical methodology or nonparametric techniques.

Of course in some situations it may well be that not enough independent blocks can be simulated to be able to obtain replications. The utility of the proposed procedure

then is that even though the user is forced back onto the regular regenerative estimate $re(n)$ he is not doing so blindly. Moreover, the single regression adjusted regenerative estimate $rare(n)$ provides him with an essentially bias free alternative to the regenerative estimate.

In Section 6 examples of the use of the protocol and its graphical displays are given for three queueing models and in Section 7 simulation results on the relative properties of $re(nm)$ and $arare(m,n)$ are presented. Section 8 contains a summary, indicates related topics for further research, and gives examples of situations other than regenerative simulations for which this methodology applies. Table 4 gives a summary of the notation used in the paper.

3. THE REGRESSION ADJUSTED REGENERATIVE ESTIMATE

Under broad conditions the expected value $E(re(n))$ has the form

$$E(re(n)) = \beta_0 + \beta_1 n^{-1} + \beta_2 n^{-2} + \cdots, \tag{3.1}$$

where $\beta_0 = E(W)$. Let a section of n regenerative blocks be broken up into subsections of lengths n_k for $k = 1,\ldots,K$ and let $m_k = [n/n_k]$, where $[x]$ denotes the greatest integer less than or equal to x. For each k, m_k regenerative estimates, $re(j,n_k)$, $j = 1,\ldots,m_k$, are formed from disjoint blocks of n_k cycles.

Regression adjusted regenerative estimates are formed by estimating β_0 in (3.1) using the $re(j,n_k)$'s for $j = 1,\ldots,m_k$, $k = 1,\ldots, K$, and the equation (3.1) to some degree $d = 1,2$, or 3. This can be done in three distinct ways:

(i) By forming average regenerative estimates, $are(m_k,n_k)$, for $k = 1,\ldots,K$ where $are(m_k,n_k) = \sum_{j=1}^{m_k} re(j,n_k)/m_k$, and using these as dependent variables in the regression on $\beta_0 + \beta_1 n_k^{-1} + \cdots + \beta_d n_k^{-d}$. The averaging will tend to make the dependent variables more normal than the individual $re(j,n_k)$'s and since $var(re(j,n_k)) \sim \gamma/n_k + \delta/n_k^{3/2} + \cdots$, then if we neglect the higher-order terms $var(are(m_k,n_k))$ is approximately γ/n for all $k = 1,\ldots,K$. Thus an unweighted least squares estimate of β_0 will produce $\hat{\beta}_0$, the $rare(n)$ estimate unbiased out to terms up to $1/n^d$.

(ii) By using the $re(j,n_k)/m_k$ for $j = 1,\ldots,m_k$, $k = 1,\ldots,K$ in a weighted regression on $\beta_0/m_k + (\beta_1/m_k)n_k^{-1} + \cdots + (\beta_d/m_k)n_k^{-d}$. This is equivalent to (i) if least squares estimation of $\beta_0,\ldots,\beta_d$ is performed.

(iii) Since it will be seen in later sections that $re(j,n_k)$ is highly non-normal for small n_k, robust regression methods could be used to eliminate the influence of outliers. However, since the $re(j,n_k)$'s are generally not symmetric random variables, the $\hat{\beta}_0$ estimate will generally be biased.

In this paper the *rare* estimate is always β_0 obtained as in (i), which like $re(n)$ is a statistic involving $(Y_1,\ldots,Y_n)$ and $(\tau_1,\ldots,\tau_n)$. The use of robust regression techniques will be described elsewhere.

There are still three problems which make the regression non-standard:

(i) How large should K be, given that it must be at least greater than d+1?

(ii) The $are(m_k,n_k)$ are correlated, which seems to mitigate against making K too large.

(iii) Given that K is chosen, how should the n_k be chosen? This is an experimental design problem.

These are difficult and comprehensive analytic problems, not considered here. However, it is possible to show that the correlation between $are(m_k,n_k)$ and $are(m_j,n_j)$ converges to one if $n \to \infty$ in such a way that m_k and m_j are fixed divisors of n. Furthermore in this case $var(rare(n))/var(re(n))$ also converges to one. For small to medium sample sizes one suspects that the correlation will inflate the variance of the $rare(n)$ estimate as compared to the variance of the $re(n)$ estimate. This frequently happens in small samples with the jackknife. However, simulation studies presented in Section 7 show that $var(rare(n))$ is only very slightly larger than $var(re(n))$ and that this result is robust with respect to the choice of K and the n_k's. Of course, choosing n_k too small will result in non-normality and, possibly, bias (the degree, d, being inadequate).

We emphasize here that the preference for the $rare(n)$ estimate over and above a jackknife estimate or a Tin estimate, is that its components allow one to

study the evolution of the regenerative estimate. The graphics to do this are discussed in the next section.

Of course one utility of the jackknife is that it can be used with small samples that a jackknifed variance estimate is available. However, this last advantage is not an issue here since a standard variance estimate is available for regenerative estimate $re(n)$; furthermore, once a sectioning into m sections is achieved, even this can be dispensed with. A regression-based estimate of the variance of $re(n)$ can be constructed but is not discussed here.

4. DESCRIPTION OF THE GRAPHICS

Three different types of graphs are used to display and analyze the output. The first two, the Basic Graph and the Retrenched Graph, are quite similar and are concerned with identifying the evolution of bias, skewness and departures from normality (in the form of individual outliers or outlying sample paths) in the regenerative and *rare* estimates. The third graph, the running normal graph, specifically displays the convergence to normality of the estimates. These three types of graphs, described in this section, all deal with the data of primary concern, namely, the *re*, *are*, *rare*, and *arare* estimates. They are novel and are adaptations of ideas which are pertinent in other estimation contexts.

(Other plots that may be of interest in a regenerative simulation include histograms and box plots of $\{Y_k, k \geq 1\}$ and $\{\tau_k, k \geq 1\}$. Scatter plots of $\{(Y_k, \tau_k), k \geq 1\}$ show the high correlation that typically exists between Y_k and τ_k. These last three plots are useful in explaining the skewed distribution often exhibited by regenerative estimates. However, since we feel their importance is secondary and they are well known they will not be discussed here.)

We proceed by first describing the box plot which is a primary component of the Basic and Retrenched Graphs. A description and interpretation of the Basic and Retrenched Graphs comes next, followed by a description of normal plots and the third type of graph, a sequence of normal plots called a running normal graph. A protocol for the interactive use of these graphs is outlined in Section 5 and

examples of its use is given in Section 6. The actual plots in this paper were produced on a Tektronix 4013 terminal using an APL graphics package. The user may wish to refer to Table 4 for a summary of the notation used in the paper.

4.1. The Box Plot

A convenient and compact technique for displaying the distribution of a batch of data is the box plot. These plots were introduced by Tukey (1977), although we prefer the form adapted by McNeil (1977). The box plot is an excellent tool for identifying skewness and asymmetry in the data as well as for detecting outliers. The plot is therefore well suited to analyzing the often highly skewed and sometimes wild output of queueing simulations (see, e.g., the left-most box plot of Figure 4C at $n = 50$; the 160 regenerative estimates which make up the box plot are highly positively skewed and clearly non-normal).

Figure 1 shows a sample box plot. The body of the box plot displays the median and the lower- and upper-quartiles of the data. Let q represent the (estimated) interquartile distance; q = upper quartile minus lower quartile. Those data points between the lower quartile minus 1.5q and the lower quartile minus q are marked by light circles, as are those data points between the upper quartile plus q and the upper quartile plus 1.5q. Those values below the lower quartile minus 1.5q or above the upper quartile plus 1.5q are marked by dark circles and are meant to be indicated as outliers. For normally distributed data, approximately 5% of the points should be marked by light circles, but only about 1 in 200 should be marked by dark circles. One problem with box plots is that they suppress sample size and this can lead to erroneous inferences; see McGill, Tukey and Larsen (1978) for extended box plots which take account of this. For samples of size less than 9 the box plots are not very meaningful and the data points are laid out vertically by magnitude for visual inspection.

4.2. The Basic and Retrenched Graphs

The two primary graphs used in the output analysis are the Basic Graphic and the Retrenched Graph. The Basic Graph provides, in the first stage of the protocol,

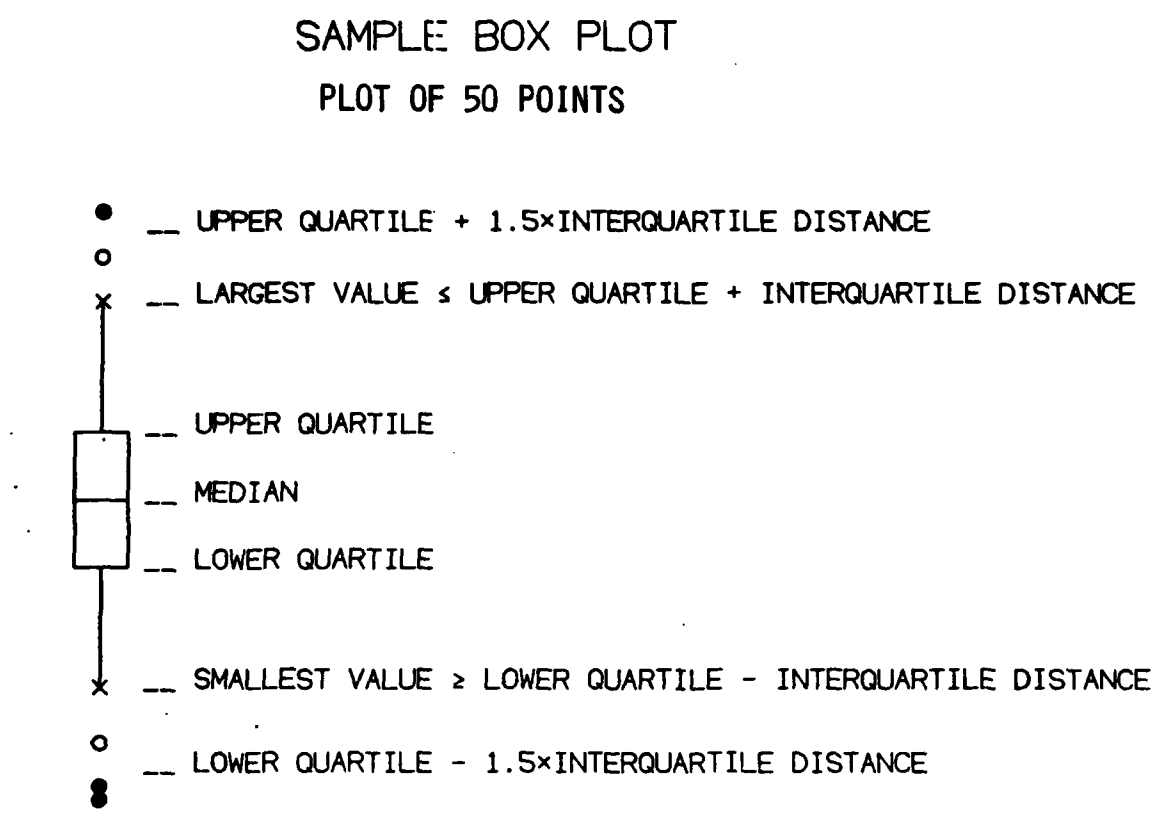


FIGURE 1. Example box plot for fifty data points from a regenerative simulation. The interquartile distance equals the estimated upper quartile minus the estimated lower quartile. The light circles are data points which fall between the largest value less than or equal to the upper quartile plus the interquartile distance and the upper quartile plus 1.5 times the interquartile distance. The dark circles are data with values above this latter point. Similarly for the lower part of the box plot.

preliminary estimates of the section length, n, and the degree of the polynomial, d, needed in the regression to reduce the bias to an inconsequential level. As the total run length increases, the Retrenched Graph is used to adjust these quantities and to monitor the distributional convergence of the *re*'s, *are*'s and *rare*'s.

The Retrenched Graph displays the results of performing m i.i.d. replications of the regression as described in Section 3. The Basic Graph is essentially a special case of the Retrenched Graph with m = 1. Examples of Basic Graphs are Figures 3A, 3B, 4A and 5A while Figures 3C, 3D, 3F, 3G, 3I, 4B, 4C, 4D, 4E, 4G and 5B are all Retrenched Graphs. The following discussion describes the features of the Retrenched Graph; it applies as well to the Basic Graph with m set equal to one. Following the description is a discussion on interpreting the results of these graphs.

Let the total run length of N cylces be divided into m replications of n cycles per replication (or section). Let the subsection lengths within each section be $\{n_k, k = 1,\ldots,K\}$ and let m_k be the number of regenerative estimates, $re(j,n_k)$, formed from n_k cycles, that may be obtained from a total of n cycles; then $m_k = [n/n_k]$.

In the Retrenched Graph (see e.g. Figure 3C), the x-axis runs from 0 to n cycles. Additional positions along the axis are provided for

(i) the single $re(mn)$, the regenerative estimate from the whole sample, marked by * and labelled REG on the right-hand side of the graph;

(ii) the single $arare(m,n)$ estimate, labelled ARARE and marked by o, with a horizontal line through it and

(iii) the box plot at position ARARE of the *rare* estimates.

Let $\overline{are}(mm_k,n_k)$ be the average of the $are(m_k,n_k)$'s, i.e. $are(mm_k,n_k) = \sum_{j=1}^{m} are(j,m_k,n_k)/m$ where $are(j,m_k,n_k)$ is the realization of $are(m_k,n_k)$ on the jth replication. The set of points $\{n_k, are(mm_k,n_k)), k = 1,\ldots,K\}$ are denoted by stars; these are unbiased point estimates of $\{E(re(n_k)), k = 1,\ldots,K\}$. The average bias curve, i.e., the set of points $\{(n_k, \sum_{i=0}^{d} \bar{\beta}_i n_k^{-i}), k = 1,\ldots,K\}$ where d is the degree of fit and $\bar{\beta}_i$ is the average of the m i.i.d. estimates of the coefficient of n^{-i} in

the bias expansion, is plotted as a connected curve. The horizontal line at height $arare(m,n)$ represents the asymptote of the average regression curve. A box plot of either $\{are(j,m_k,n_k),\ j = 1,\ldots,m\}$ (as in Figure 4G) or $\{re(j,n_k),\ j = 1,\ldots,mm_k\}$ (as in Figure 4C) is displayed at x-axis position n_k for each k. If less than nine points are available for a box plot, the individual values are plotted as magnitude-ordered circles rather than displaying a box plot for so few points; it is simple to see the spread and location of the estimates this way.

A count of the number of $rare(j,n)$ estimates which are less than $re(nm)$ is given: this gives a rough sign-test for lack of bias in the $\{rare(j,n)\}$, with $re(mm)$ being used as the true value. The values of $arare(m,n)$, $srare\ (m,n)$, $re(nm)$ and $\hat{\sigma}(nm)$, the usual regenerative standard deviation estimate based on all nm cycles, are printed below the x-axis. In the BASIC GRAPH when $m = 1$ the *ARARE* standard deviation is set equal to zero; the intent is to eventually use here a regression based estimate for the standard deviation of $re(n)$:

With regard to interpretation, the ideal is a Retrenched Graph for which

(i) the average bias curve is smooth and, beyond [n/2], close to its asymptote and relatively flat ;

(ii) the estimates $re(nm)$, $arare(m,n)$ and $\overline{are}(m,n)$ do not differ substantially;

(iii) the distributions of $\{re(j,n),\ j = 1,\ldots,m\}$ and $\{rare(j,n),\ j = 1,\ldots,m\}$ are symmetric about $re(nm)$ and their respective means, and

(iv) $srare(m,n)$ and $\hat{\sigma}(nm)$ are approximately equal.

Figure 4G is just such a graph. When n is small enough that $re(n)$ is still biased and $rare(n)$ needs to be used as a less biased estimated, this bias shift will show clearly in the relative shift of the box plots of the $re(j,n)$'s and the $rare(j,n)$'s as in Figure 4C.

Properties (i) and (ii) listed above and illustrated in Figure 3B are the ideals in a Basic Graph. Because of the high correlation between $are(2,[n/2])$ and $re(n)$, a flat bias curve beyond $[n/2]$ indicates that the difference between $E(re(n))$ and

$E(re([n/2]))$ is small. This implies that the bias in $re(n)$ is also small. The Basic Graph is therefore very useful in determining the section length, n, to be replicated even though the variability in the estimates may still be substantial. Since a Basic Graph consists of a single replication, only limited distributional information can be learned from it. However, the distributions of $\{re(j,n_k), j = 1,\ldots,m_k\}$ will give the user a general idea as to the variability in the data.

Further interpretation of these graphs is given in the protocol of Section 5 and the examples of Section 6.

4.3. Normal Plots and the Running Normal Plot

To detect a normal distribution, a more detailed analysis of the data than is possible with box plots is accomplished with normal plots. A number of statistical tests accompanying the normal plots may be used to formally test the hypothesis of normality. Normal plots [see e.g. Wilk and Gnanadesikan (1968)] are formed as follows.

Let $\underset{\sim}{X} = \{X_1, X_2, \ldots, X_n\}$ be an i.i.d. sequence of random variables with continuous distribution function $F(x)$, $-\infty < x < \infty$. Let $\Phi(x)$, $-\infty < x < \infty$, be the standard normal distribution function and let $\Phi^{-1}(\cdot)$ and $F^{-1}(\cdot)$ be the inverse distribution functions of Φ and F respectively. Let $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ be the order statistics of $\underset{\sim}{X}$. An estimate of $F^{-1}(k/n+1)$ is $X_{(k)}$ and for any constants a, b $(b > 0)$ a plot of the pairs of points $\{(\Phi^{-1}(k/n+1), (X_{(k)}-a)/b), k = 1,\ldots, n\}$ will, for large n, result in a straight line if $\underset{\sim}{X}$ is a set of normally distributed random variables (i.e. $F = \Phi$). We standardize the plot by setting $a = \bar{X} = \sum_{k=1}^{n} X_k/n$ and $b^2 = S^2 = \sum_{k=1}^{n} (X_k-\bar{X})^2/(n-1)$. Departure from linearity in this normal plot indicates non-normality.

Formal tests of normality that we use include the Shapiro-Wilk statistic (1965) which was also used by Fishman (1977). For ease of computation we apply the test to at most 50 of the points. The significance level of the test was computed using the approximation in Shapiro and Wilk [1968]. Additional test statistics are the coefficients of skewness and kurtosis (Snedecor and Cochran (1967), p. 86) estimated by $\mu_3/\sqrt{\mu_2^3}$ and $(\mu_4/\mu_2^2) - 3$, respectively, where $\mu_j = \sum_{k=1}^{n} (X_k-\bar{X})^j/n$.

If the X_k's are normally distributed, then for large n these moment statistics are approximately normally distributed with means zero and standard deviations $\sqrt{6/n}$ and $\sqrt{24/n}$ respectively. A coefficient of skewness or kurtosis that differs from zero by more than about two times its standard deviation indicates significant departure from normality. Older tests of normality were based on these sample coefficients; recently interest has renewed is using the pair of statistics as a test for normality (see Shapiro and Wilk (1965)). One can also use Kolmogorov-Smirnov type statistics suitably adapted for the fact that the mean and standard deviations are estimated from data (Lilliefors (1967)).

Figures 3E, 3H, 3J, 4F, and 4H are examples of a sequence of normal plots of simulation generated estimates; we call this a running normal plot. The goal of these graphs is to monitor the convergence to normality of the regenerative estimates $re(j,n_k)$ and to determine at what point normality may safely be assumed to exist. In addition to the features previously mentioned, a line through (0,0) with slope one is drawn for each normal plot in the sequence. For any subsection length, n_k, either $\{re(j,n_k),\ j = 1,\ldots,m_k m\}$ (as in Figure 4F and labeled REG) or $\{are(j,n_k),\ j = 1,\ldots,m\}$ (as in Figure 4H and labeled AREG) may be plotted in addition to a plot of $\{rare(j,n)\ j = 1,\ldots,m\}$.

Placing multiple normal plots on a single running normal graph is very useful in identifying trends in convergence. For example, in Figure 4F the plots of $\{re(j,500)\}$, $\{re(j,1000)\}$ and $\{rare(j,1000)\}$ when examined individually do not indicate significant departures from normality. However, when placed in sequence along with the plot of $\{re(j,250)\}$, they are seen to exhibit a pattern of skewness and nonnormality similar to that of $\{re(j,250)\}$ (although to a lesser extent). Since normal plots and tests are difficult to interpret for small data sets, the comparison between normal plots at small and large values of n is helpful.

5. A PROTOCOL FOR SEQUENTIAL APPLICATION OF THE GRAPHICS

The following protocol outlines the sequential use of the regression adjusted regenerative estimates and graphics in analyzing the output of a simulation. The objective is to determine the total run length needed to achieve a given precision in the simulation and to obtain a sectioning of that run length into m replications so that the resulting *rare* estimates are approximately unbiased and normally distributed. The protocol consists of two stages. The first stage of the sequential protocol provides a preliminary estimate of the section length and the degree of the fitted regression curve which is needed to eliminate the concern over bias, i.e. to obtain an approximately bias-free *rare* estimate. It also provides an initial estimate of the simulation's variability. The second stage of the protocol increases the run length until the desired accuracy is achieved. The degree and section length are adjusted along the way to maintain low bias and increase distributional symmetry. When the final accuracy is achieved, further adjustments are made to gain normality in the estimates from each section.

We suggest that between 10 and 20 replications be available at the end of the simulation, since an adequate estimate of the variance requires at least 10 observations (replications). Twenty observations should be sufficient for most comparison purposes (Mosteller and Tukey, 1977, Ch. 7) and keeping more than 20 sections will only slow convergence to normality. Thus, for example, if the plots showed that the required precision was obtained with 40 sections, each of length 100 cycles, and that the *rare*(j,100)'s were approximately normal, we would still recommend refiguring the simulation to 20 sections, each of length 200 cycles.

The protocol is intended only as a guide and a user need not adhere to it strictly. The whole point of the graphics is to assist users in exercising their own judgment.

Throughout the protocol we use N to represent the total simulation run length (in cycles). The section length, or number of cycles per replication, is

denoted by n and m is the number of sections, or replications; thus $M = [N/n]$. The subsection lengths, picked on the basis of empirical tests such as those detailed in Section 7, are defined by $n_k = [n/k]$ for $k = 1, 2, 3, 4, 7, 10$.

I. Stage I-Pilot run

1. Set N_0 = initial run length.
2. Simulate a total of N_0 cycles.
   a. Set m = number of replications = 1
   b. Set n = section length = number of cycles/replications = N_0
   c. Set subsection lengths $n_k = [n/k]$, $k = 1, 2, 3, 4, 5, 7, 10$.
3. Form Basic Graphs for degrees $d = 1,2,3$.
   a. Pick smallest d so that regression curve is smooth and fits data well. If no such d can be determined increase N_0 and go back to I.2.
   b. Find n^*, the smallest n_k so that the regression curve is nearly flat beyond n^* and there is little difference between $are([n/n^*],n^*)$, $rare(n)$ and $are([n/n'],n')$ for $n' \geq n^*$. If no such n^* can be determined increase N_0 and go back to I.2.
   c. Based on the spread of points in the Basic Graph and the regenerative standard deviation estimate, $\hat{\sigma}(N_0)$ project a new estimated run length N to yield desired accuracy.
   d. Set $n = n^*$ and go to Stage II.

(Note that one may conclude at this point that the regenerative cycles are so long that for a given precision one must accept a biased and/or non-normal regenerative estimate, although it would be better to accept the less biased but still non-normal regression adjusted regenerative estimate. Note too in the case where the variability of $re(N_0)$ and $rare(N_0)$ is high, the "power" of the procedures is low and the decisions made here may have to be modified in Stage II.)

II. Stage II--Retrenchment and Fine Tuning

1. Simulate a total of N cycles.

2. Set m = [N/n], where n was determined in Stage I.

3. Form Retrenched Graphs for degrees d = 1,2,3.

   a. Pick smallest d giving good fit to $\{\overline{are}(mm_k, n_k)\}$.

   b. If m > 20 and either

      i. $\overline{are}(mm)$ or $arare(m,n)$ show substantial bias as indicated by the average regression curve being far from its asymptote, or $arare(m,n)$ or $\overline{are}(m,n)$ being significantly different from $re(N)$, or

      ii. Box plots of $\{re(j,n)\}$ or $\{rare(j,n)\}$, j = 1,...,m, indicate asymmetry then double n and go back to II.2.

   c. Otherwise if $srare(m,n)$ and $\hat{\sigma}(N)$ are sufficiently small to yield the desired accuracy go on to II.4; otherwise project a new run length N to yield desired accuracy and go back to II.1.

4. If m > 20 set m = 20; if m < 10 set m = 10. Set n = [N/m].

   a. Form Retrenched Graph and adjust the degree as in II.3.a.

   b. Form normal plots of $\{rare(j,n)\}$; $\{re(j,n_k\}$ and $\{are(j,m_k,n_k)\}$ for k = 1,2,4.

   c. If normality is indicated then stop.

   d. If normality is not indicated, then

      i. if m = 10 increase N if possible and go to II.1 ;

      ii. otherwise reduce m, keeping $m \geq 10$, and go to II.4.

Note that in this Stage we are trying (i) to fix more accurately the sample size needed to attain the desired precision in the final estimate; (ii) to section so as to get enough degrees of freedom(m-1) in the variance estimate, $srare^2(m,n)$, so as not to sacrifice precision in this estimate, and (iii) to keep n

as large as possible so as to ensure that the *rare*(j,n)'s are approximately normally distributed. The decision may, however, be that these conflicting objectives cannot be met and a specific compromise has been adopted. This case will be illustrated in the third example in the next section.

6. SOME EXAMPLE QUEUEING SYSTEM SIMULATIONS

In this section we give examples of the use of the protocol and the associated graphics and estimates in the analysis of the output of some simple queueing systems. The systems are two single server queues with high traffic intensities and a closed queueing network model of a computer system. The single server queues, the M/M/1 and M/G/1 queues with traffic intensity 0.90 [see e.g. Kleinrock, 1975], though simple in structure, exhibit substantial bias, produce highly skewed output and require very long run lengths before stabilizing. They therefore provide challenging tests of the methodology.

The closed queueing network on the other hand, requires relatively few cycles before stabilizing. This example shows that the graphics can be of use even when few cycles are needed. However, when the number of cycles is so small as to preclude use of the Retrenched Graph, the assumptions of low bias and normality cannot be verified and caution should be exercised. In fact the conclusion may be the precautionary one that the regenerative estimate is very likely biased and non-normal, the non-normality applying also to the regression adjusted regenerative estimate.

a. The M/G/1 Queue (Examples 1 and 2)

Let $\{W_n, n \geq 0\}$ be the waiting time of the nth customer in a single server queue and let $\{A_n, n \geq 1\}$ and $\{S_n, n \geq 0\}$ be the i.i.d. sequences of interarrival and services times respectively. The waiting time sequence is defined by $W_0 = 0$ and $W_{n+1} = (W_n + S_n - A_{n+1})^+$ for $n \geq 0$ where $x^+ = \max(x,0)$. Assume

A_n is exponentially distributed with mean $1/\lambda$ and let S_n be hyperexponentially distributed; $P(S_n > X) = 1-G(x) = p \exp(-\mu_1 x) + (1-p) \exp(-\mu_2 x)$. Then $E(S_n) = 1/\mu = p/\mu_1 + (1-p)/\mu_2$. Variance$(S_n) = \sigma_s^2 = 2\{p/\mu_1^2 + (1-p)/\mu_2^2\} - 1/\mu^2$, and the coefficient of variation of S_n is $C_s = \mu\sigma_s$. If the traffic intensity is $\rho = \lambda/\mu < 1$ then regenerations occur whenever $W_n = 0$ and the waiting time has a stationary distribution with mean $E(W) = \rho(1 + C_s^2)/\{2\mu(1-\rho)\}$ (see Kleinrock (1975)). If $p = 1$ then the queue is known as the M/M/1 queue. For the M/M/1 queue of this section (Example 1) $\lambda = 0.9$, $\mu = 1.0$, $\rho = 0.9$ and $E(W) = 9.00$, whereas for the M/G/1 queue (Example 2) $\lambda = 0.9$, $\mu_1 = 0.5$, $\mu_2 = 2.0$, $p = 0.33$, $\rho = 0.9$ and $E(W) = 13.42$.

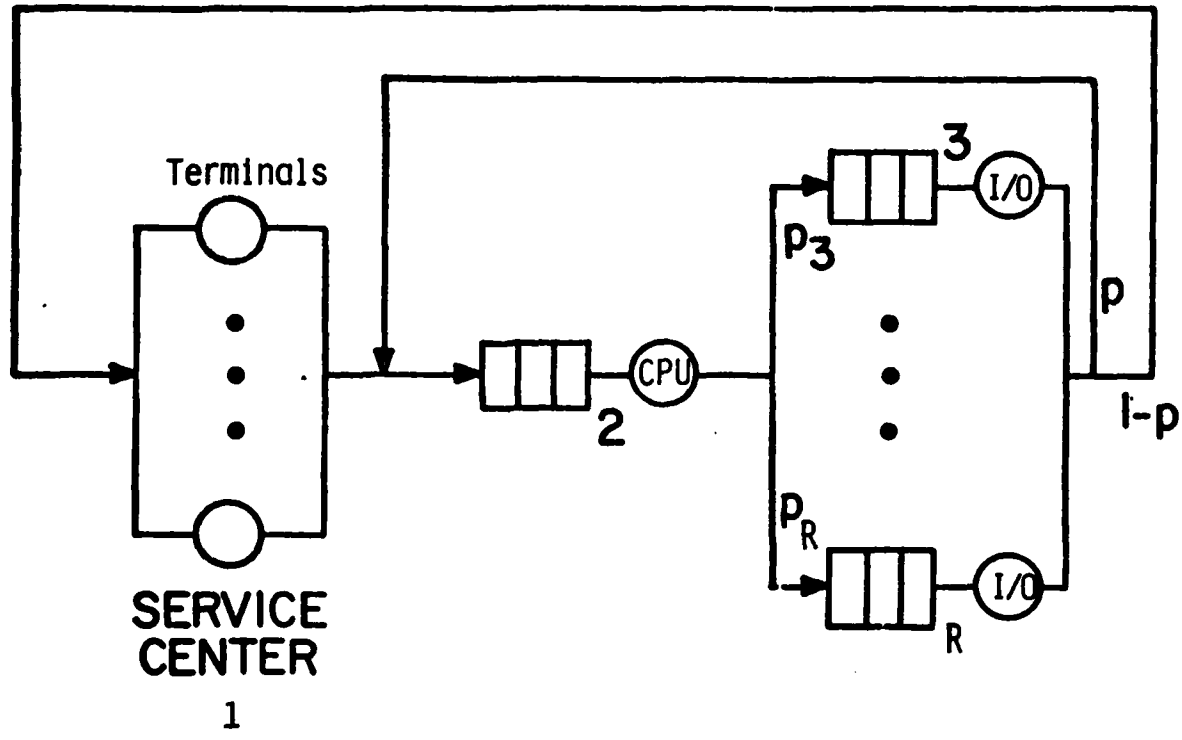
b. Closed Queueing Network (Example 3).

Consider the queueing network pictured in Figure 2. This network is a model of a timesharing computer system. There are R service centers with a fixed number, T, of customers in the network. Service Center 1 consists of T terminals. The terminal users submit jobs to a computer system consisting of a CPU (Service Center 2) and a number of peripheral input-output devices (Service Centers 3, ..., R). The CPU operates under the processor sharing discipline while the peripherial devices are each first-come, first-served single server queues. Let μ_i^{-1} be the mean of the exponentially distributed service times at service center i. Routing through the network is Markovian and the routing probabilities are given in Figure 2. By assuming that all service and routing mechanisms are mutually independent then $\underline{Q} = \{Q(t), t \geq 0\}$ is a continuous time Markov chain where $\underline{Q}(t) = (Q_1(t),\ldots,Q_M(t))$ and $Q_i(t)$ is the number of customers in service center i at time t. The assumptions on the service distributions and disciplines and on the routing are such that the equilibrium distribution of the network exists and has a product form (Baskett, et al. (1975)). Define a response time to be the time from when a customer leaves service center 1 until that customer next returns to service center 1 and let W_n be the just completed response time of the nth customer to arrive at service center 1.

Then $\underset{\sim}{W} = \{W_n, n \geq 0\}$ is regenerative with regenerations occurring at points n such that the nth customer arrives at service center 1 leaving centers 2, ... , R, empty. Again we shall be interested in the expected stationary response time E(W) which is known to be finite (Lavenberg and Sauer (1977)). Let ρ_i be the steady state utilization of service center i. The particular parameters chosen for this model are listed in Figure 2 and yield $\rho_2 = 0.894$, $\rho_i = 0.268$ for $i \geq 3$ and $E(W) = 8.65$.

The application of the methodology to these three examples is given in the Figures; each Figure has a Caption and an Interpretation which, when read sequentially describe the evolution of the simulation.

Closed Queueing Network



System Parameters

T = 10	R = 6	E(W) = 8.650
μ_1^{-1} = 10.0	μ_2^{-1} = 0.250	μ_i^{-1} = 0.300 for $i \geq 3$
p = 0.850	p_i = 0.250 for $i \geq 3$	
ρ_2 = 0.895	ρ_i = 0.268 for $i \geq 3$	

FIGURE 2. Closed queueing network which models a timesharing computer system.

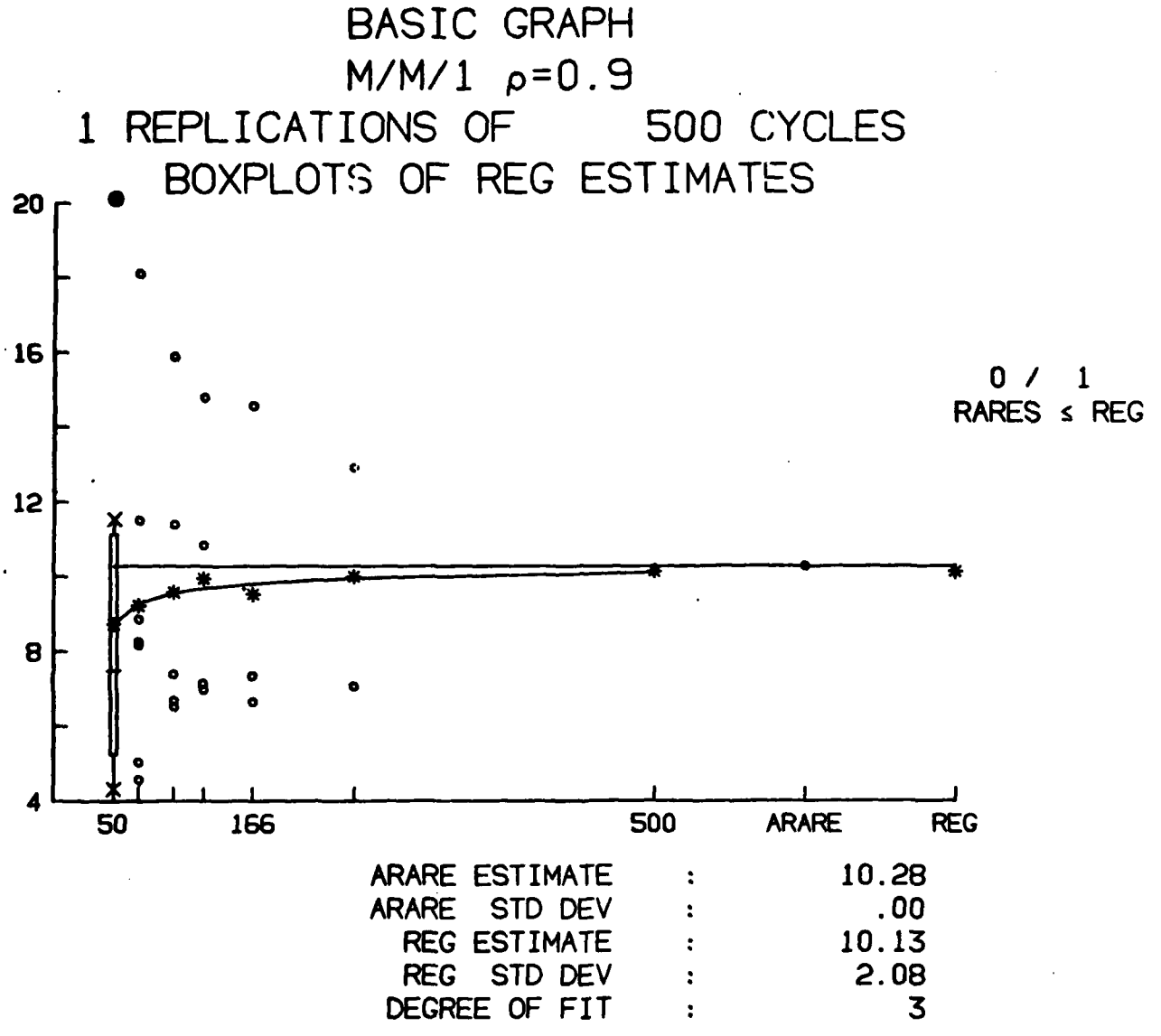


FIGURE 3A. Example 1. Stage I, steps 1 and 2. N_0 set at 500 cycles. A decision has been made on the basis of similar graphs that d = 3 is probably needed.

Interpretation. Since the precision ($2\sigma/\mu$) is about 40%, and the regression curve is still coming up to the asymptote, N_0 = 500 may be too small. Therefore return to Stage I.1 and set N_0 = 1000.

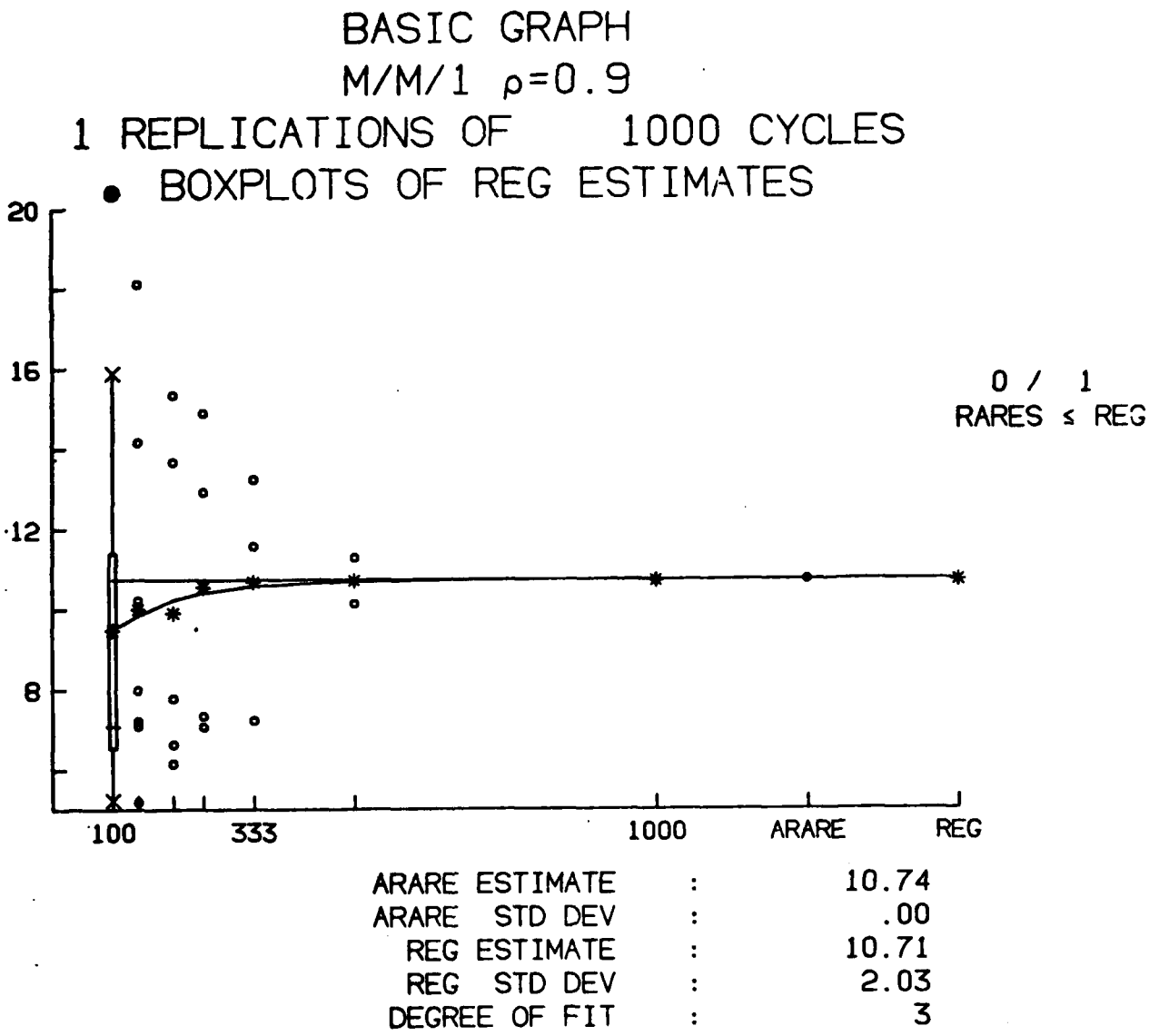


FIGURE 3B. **Example 1. Stage I, steps 1 and 2 with N_0 increased to 1000.**

Interpretation. Since the regression curve beyond 500 and *re*(1000) and *rare*(1000) are approximately equal, simulation out to n = 1000 is likely to be adequate. Since it is estimated that about 16 times as many cycles are needed to cut the standard deviation down by a quarter, an exploratory excursion to Stage II with N = 4000 is taken.

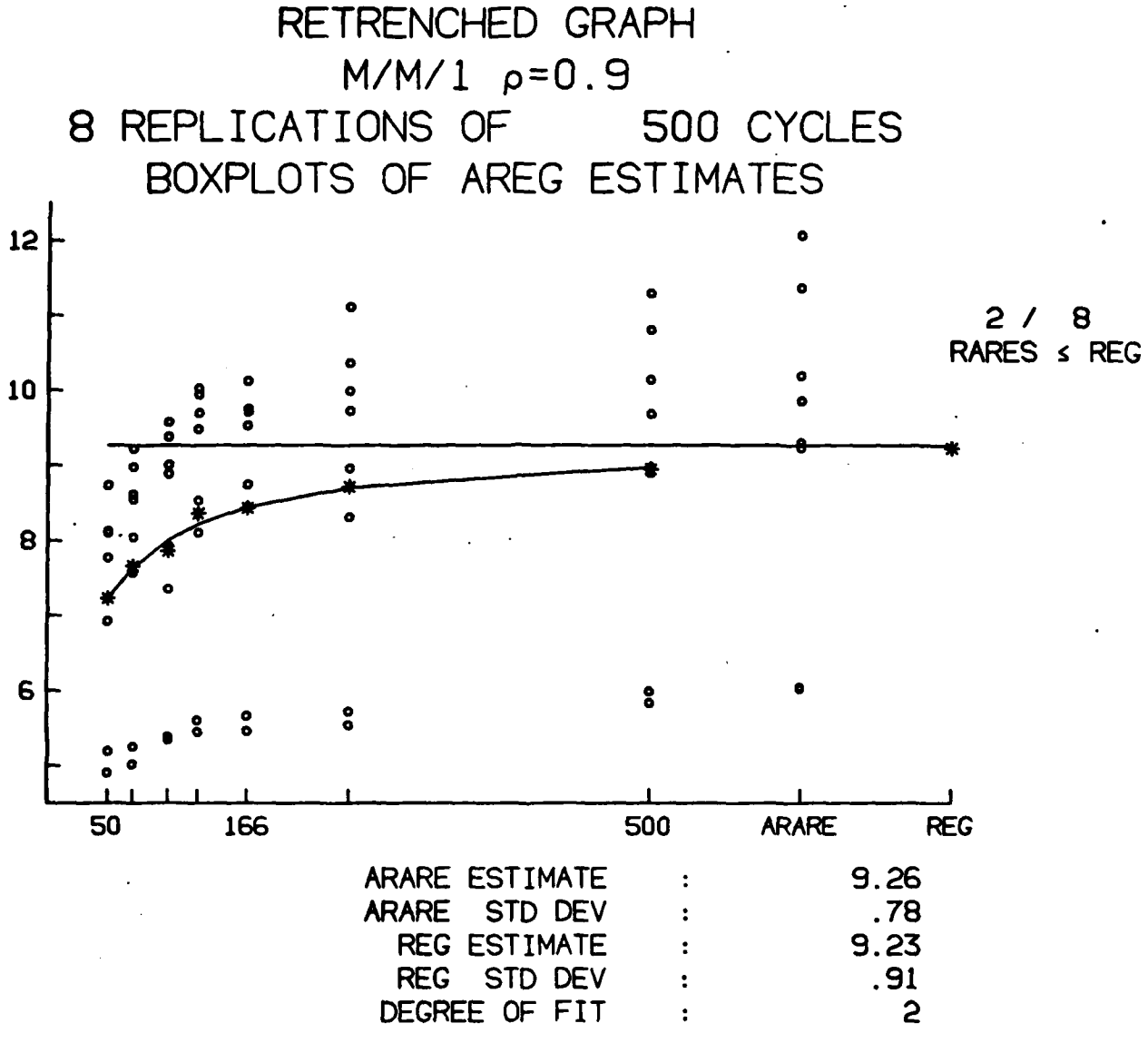


FIGURE 3C. Example 1. Stage II, step 3a. N = 4000 and d cut back to 2. Also n is cut back to 500 to determine whether, with this greater precision, a firm decision can be made on section length n. Note that the box plots are of the $are(j,n_k)$ which show less skewness then the $re(j,n_k)$ in Figure 3b.

Interpretation. The $srare(8,500)$ of 0.78 is lower than the regenerative s.d. estimate of 0.91. Clearly about two to three times as many cycles are needed. There is also still bias, as illustrated by the upward displacement of the data points at ARARE compared to the data points at n = 500. This shift is not statistically significant but is reinforced by the fact that the regression curve is increasing. The degree d = 2 is now seen to give a good fit. For illustrative purposes the next step is to repeat this Retrenched Graph with twice the number of cycles.

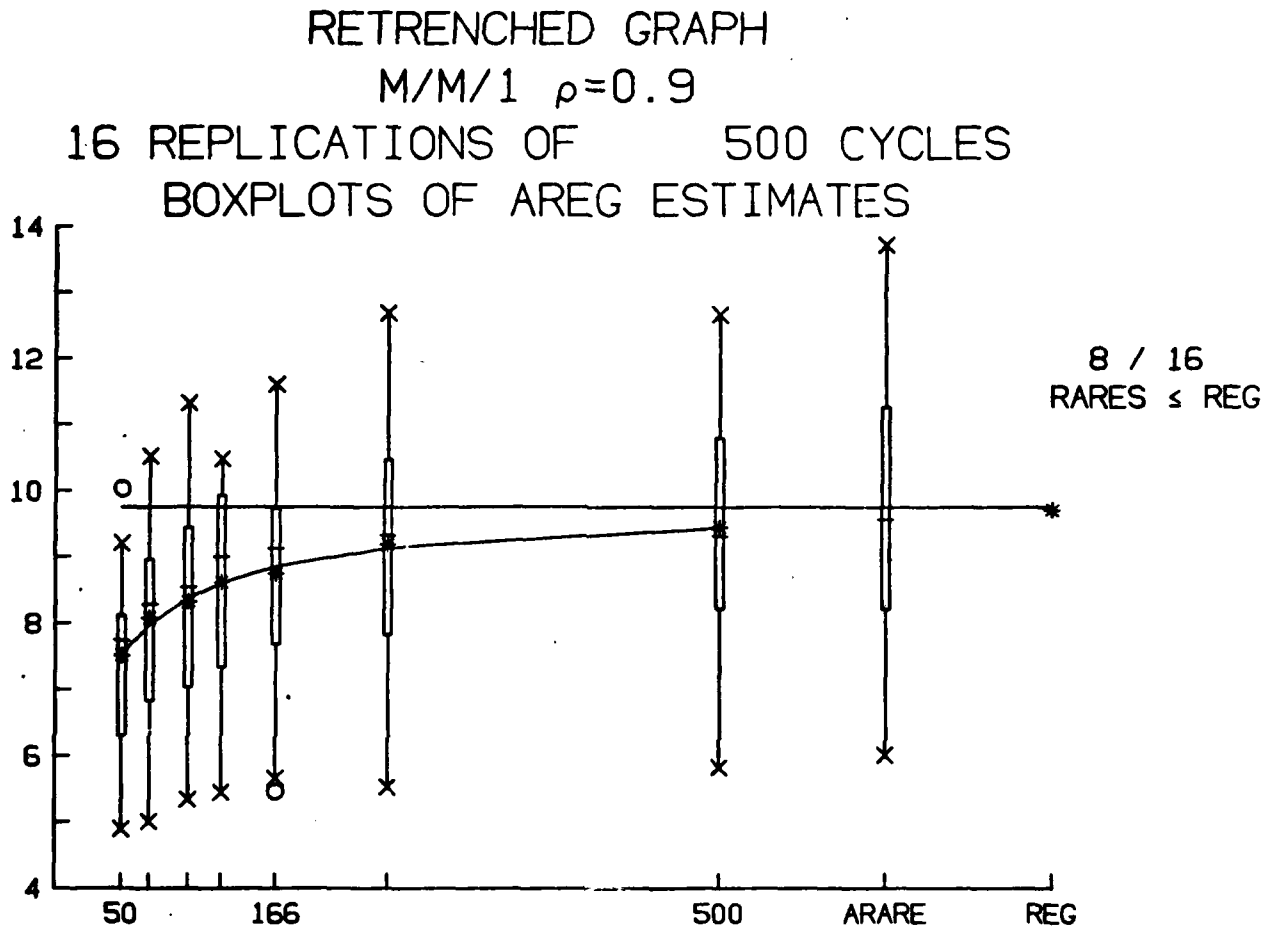
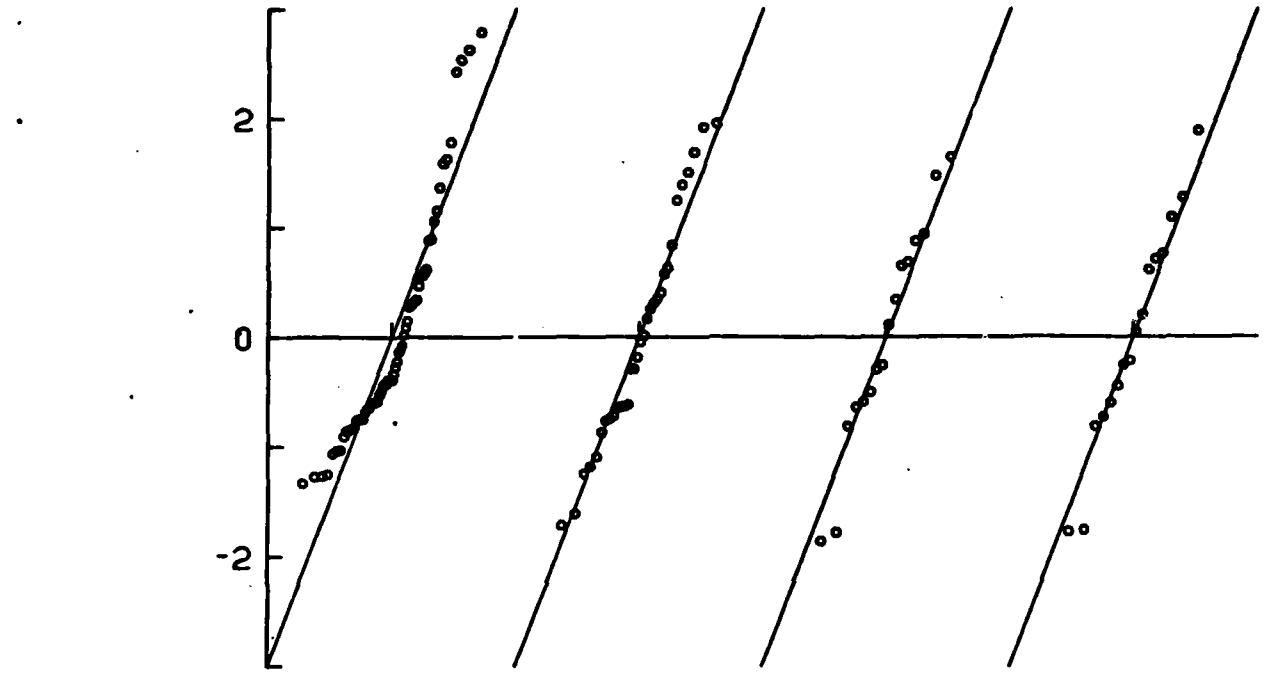


FIGURE 3D. Example 1. Stage II, step 3 repeated with N = 8000 for illustration.

Interpretation. The presence of slight bias in $re(j,500)$ is confirmed. It is also clear from the standard deviation estimates that more cycles are needed to reduce the standard deviation to less than 0.5. At this point a running normal plot is helpful and is given in the next figure.



	REG[125]	REG[250]	REG[500]	RARE	
SHAPIRO-WILK	.866(.000)	.955(.234)	.962(.673)	.974(.873)	(SIG LEV)
SKEW	1.143(.306)	.362(.433)	-.216(.612)	-.075(.612)	(SD SKEW)
KURTOSIS	.693(.612)	-.751(.866)	-.708(1.225)	-.634(1.225)	(SD KURT)

FIGURE 3E. Example 1. Supplement ot Figure 3D. N = 8000 cycles.

Interpretation. Note that while the $re(j,125)$'s are clearly non-normal, there is no indication, graphical or from the test statistics, of departure from normality either in $re(j,500)$ or $rare(j,500)$. The number of cycles will now be doubled to reach the desired precision; but first the need to go to $n = 1000$ for reasons of bias is explored.

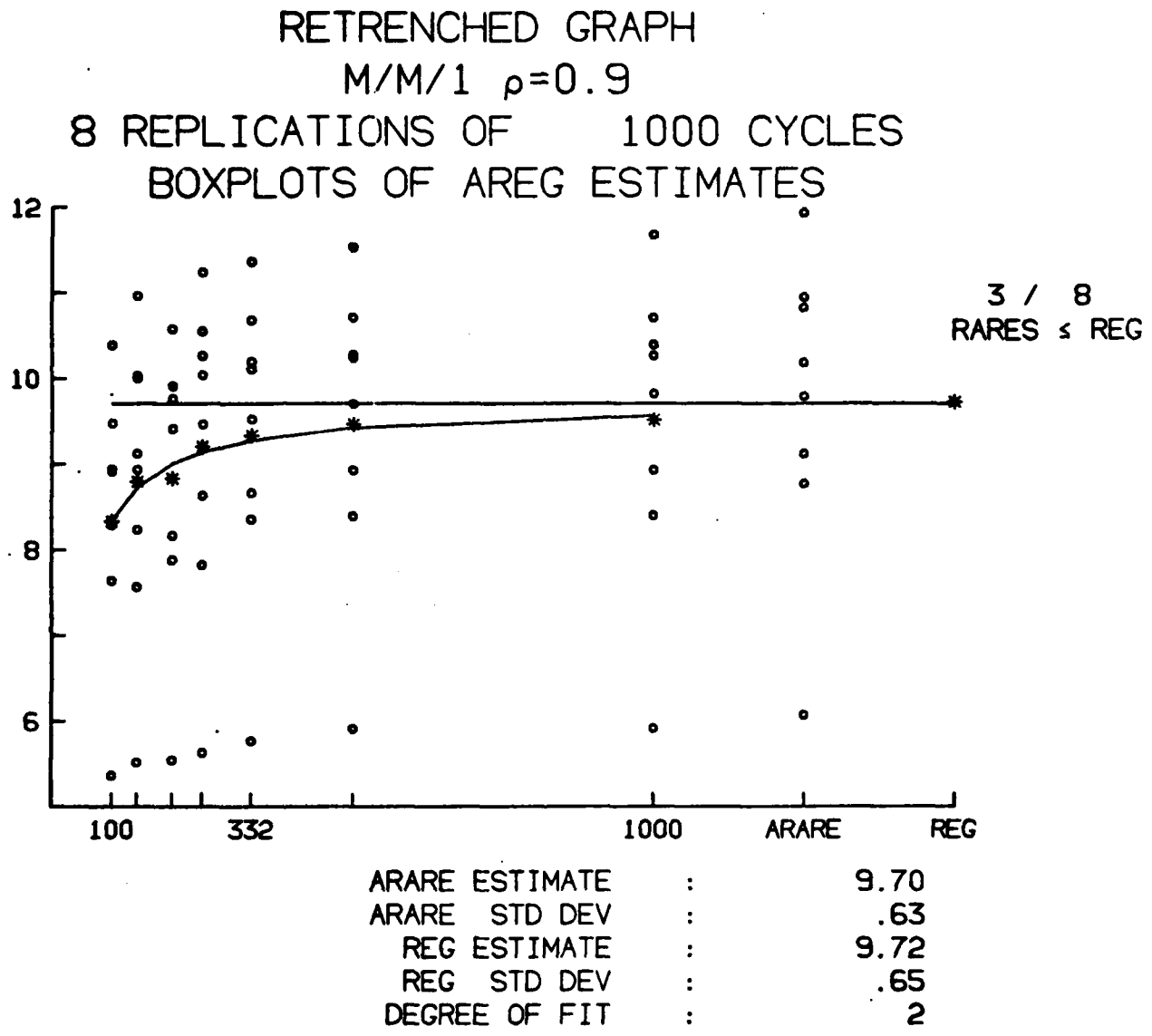


FIGURE 3F. **Example 1. N = 8000 cycles. Illustrative retrenched graph giving confirmation of the basic graph in Figure 3b.**

Interpretation. The bias, if any, at n = 1000 is small, and would be taken care of by using the *rare*(j,1000)'s instead of the *re*(j,1000)'s. Notice the evolution of the extreme, low outlier.

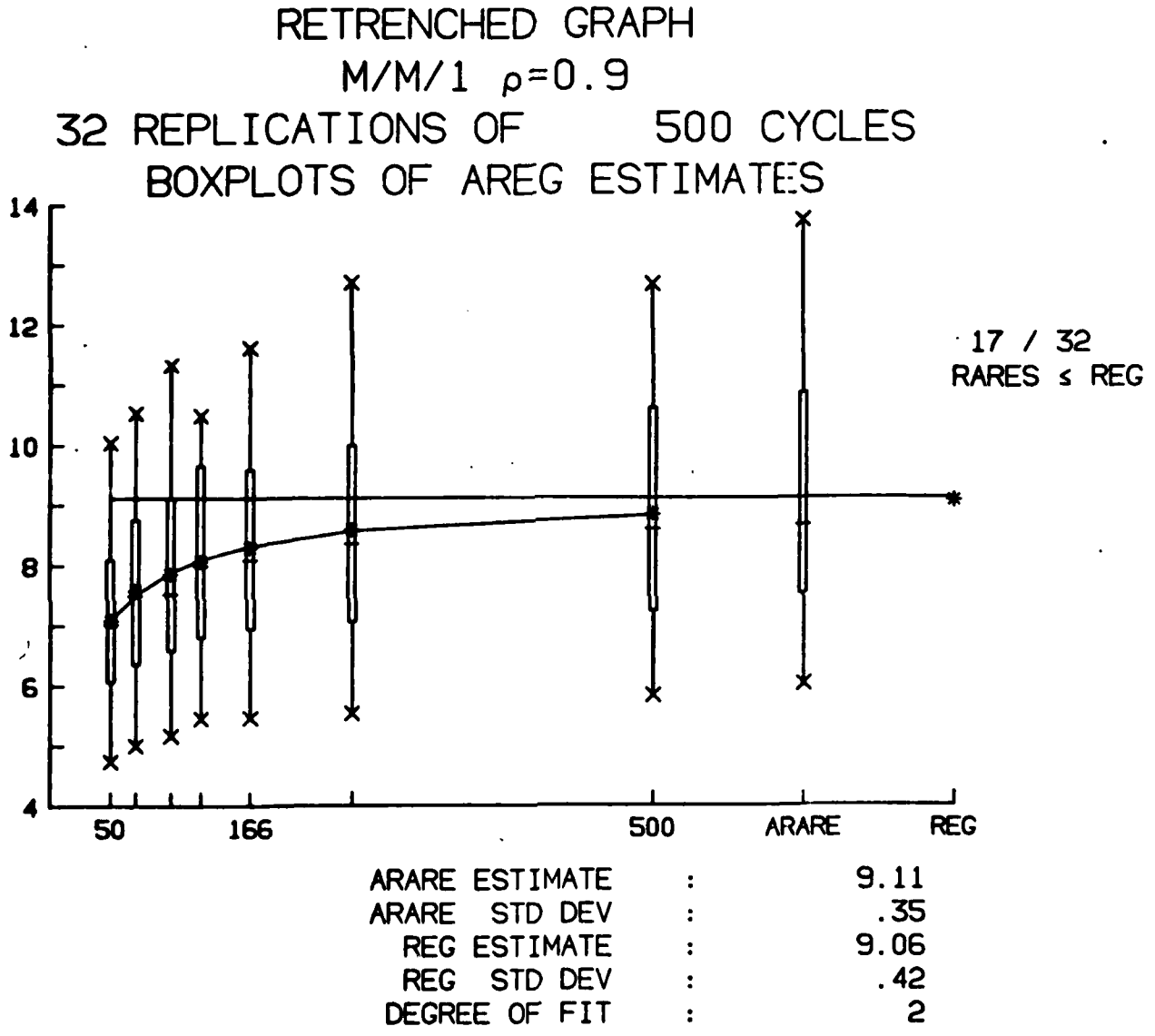
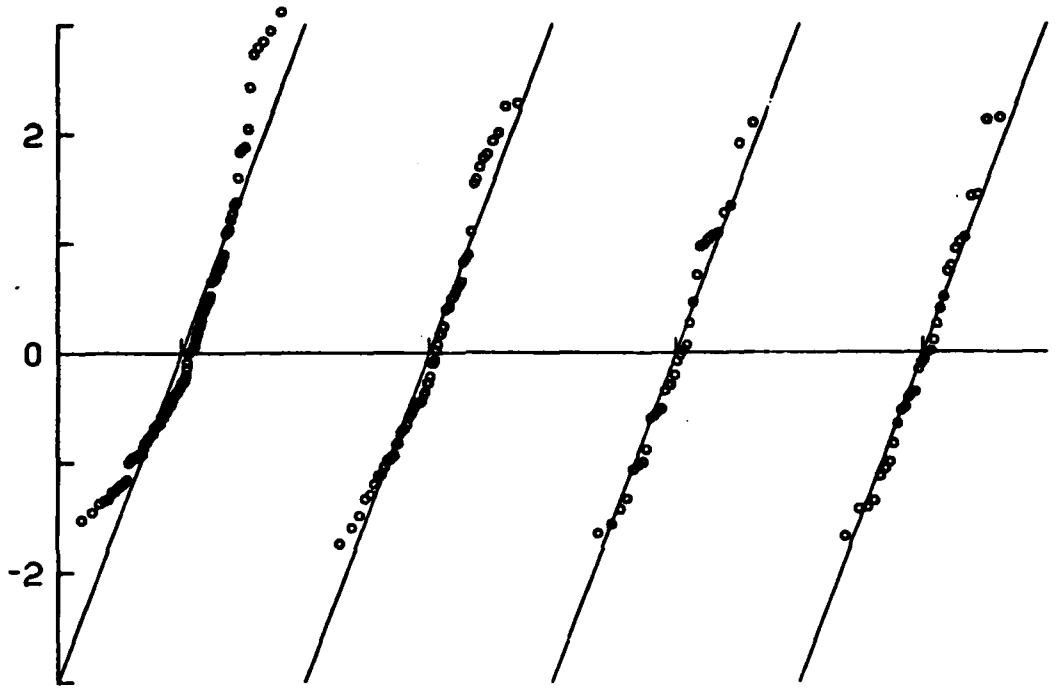


FIGURE 3C. Example 1. Stage II, step 3 with N = 1600 cycles.

Interpretation. The two estimated standard deviations are close and indicate that at least 10% precision has been attained. We are at Stage II, Step 3C here and ready to go to Stage II, Step 4. Since $m > 20$ and there is some bias, n should at least be doubled. First we look at the normality in the next figure.



	REG[125]	REG[250]	REG[500]	RARE	
SHAPIRO-WILK	.866(.000)	.942(.029)	.961(.339)	.965(.428)	(SIG LEV)
SKEW	1.124(.217)	.592(.306)	.265(.433)	.370(.433)	(SD SKEW)
KURTOSIS	1.057(.433)	-.404(.612)	-.829(.866)	-.570(.866)	(SD KURT)

FIGURE 3H. Example 1. N = 16000 cycles.

Interpretation. Another look at normality with twice the amount of data as was available in Figure 3D. Again there is no formal indication of departure from normality in the *rare*(j,500)'s, but it is to be recalled that with only 32 data points, tests for normality have relatively low power. A visual comparison of the *rare*(j,500)'s to the *re*(j,125)'s and *re*(j,250)'s indicates, however, the possibility of nonnormality. Since m = 32 > 20, n is now doubled.

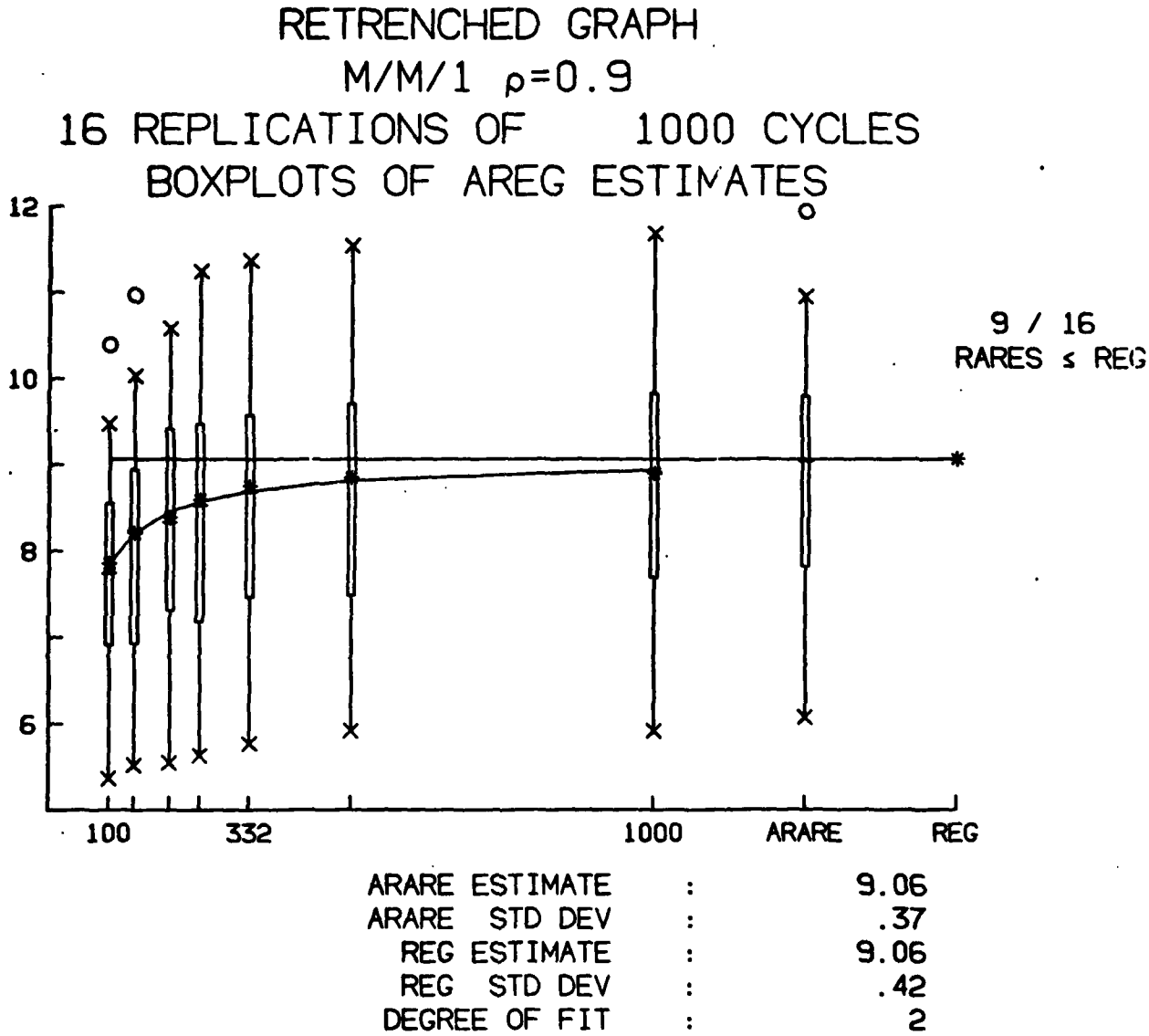
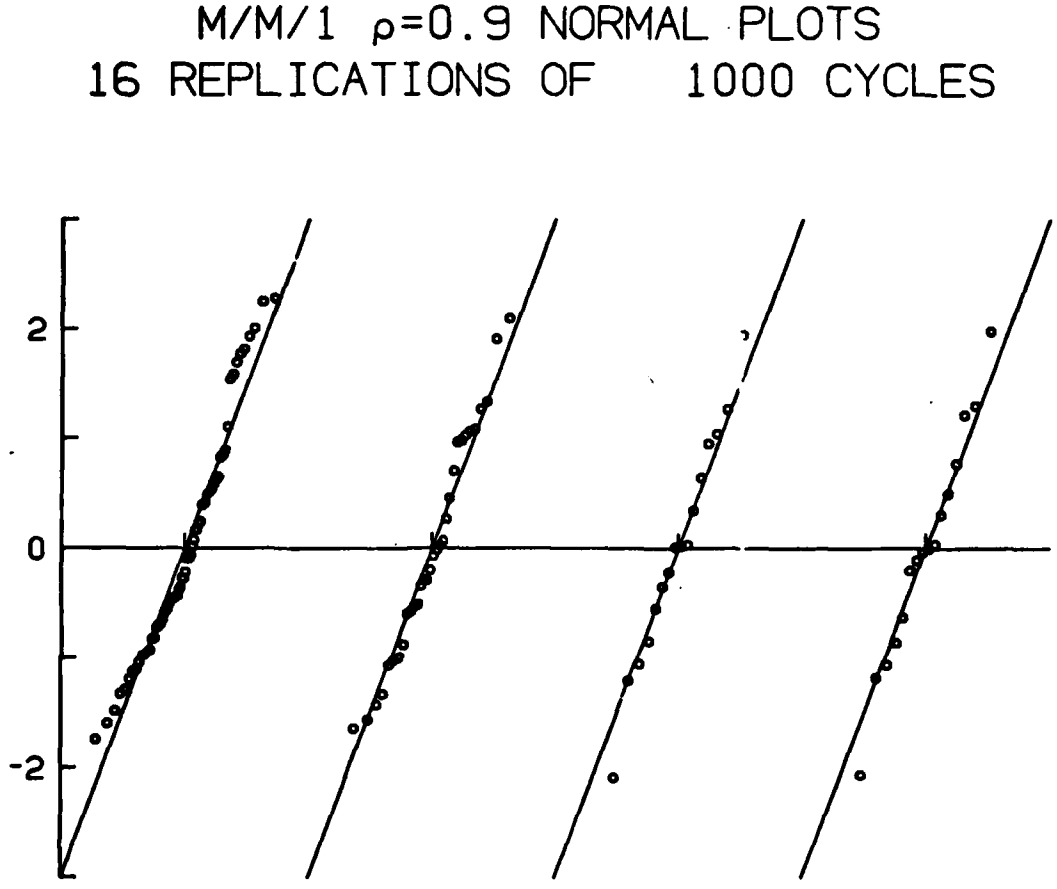


FIGURE 3I. Example 1. N = 16,000 cycles. Final steps of Stage II of the protocol. Note the similarity between this Figure and Figure 3F illustrating the stability reached with this sample size.

Interpretation. The bodies of the box plots of the *rare*(j,1000)'s and the *re*(j,1000)'s (for j = 1,...,16) indicate that there is very little to choose between them; their spreads are similar. Also there are no extreme values or outliers. The estimated standard deviation of 0.37, *srare*(16,000) indices that better than 10% precision has been achieved. A retrenched graph with degree d = 3 is indistinguishable from this graph.



	REG[250]	REG[500]	REG[1000]	RARE	
SHAPIRO-WILK	.941(.026)	.961(.339)	.992(.999)	.988(.993)	(SIG LEV
SKEW	.592(.306)	.265(.433)	-.066(.612)	-.007(.612)	(SD SKEW
KURTOSIS	-.404(.612)	-.829(.866)	-.383(1.225)	-.319(1.225)	(SD KURT

FIGURE 3J. Example 1. N = 16,000 cycles. Final look at normality Stage II, Step 4C.

Interpretation. For illustration the $re(j,n_k)$'s are shown, rather than the $are(j,n_k)$. There is no indication of departure from normality in the $rare(j,1000)$'s or the $re(j,1000)$'s.

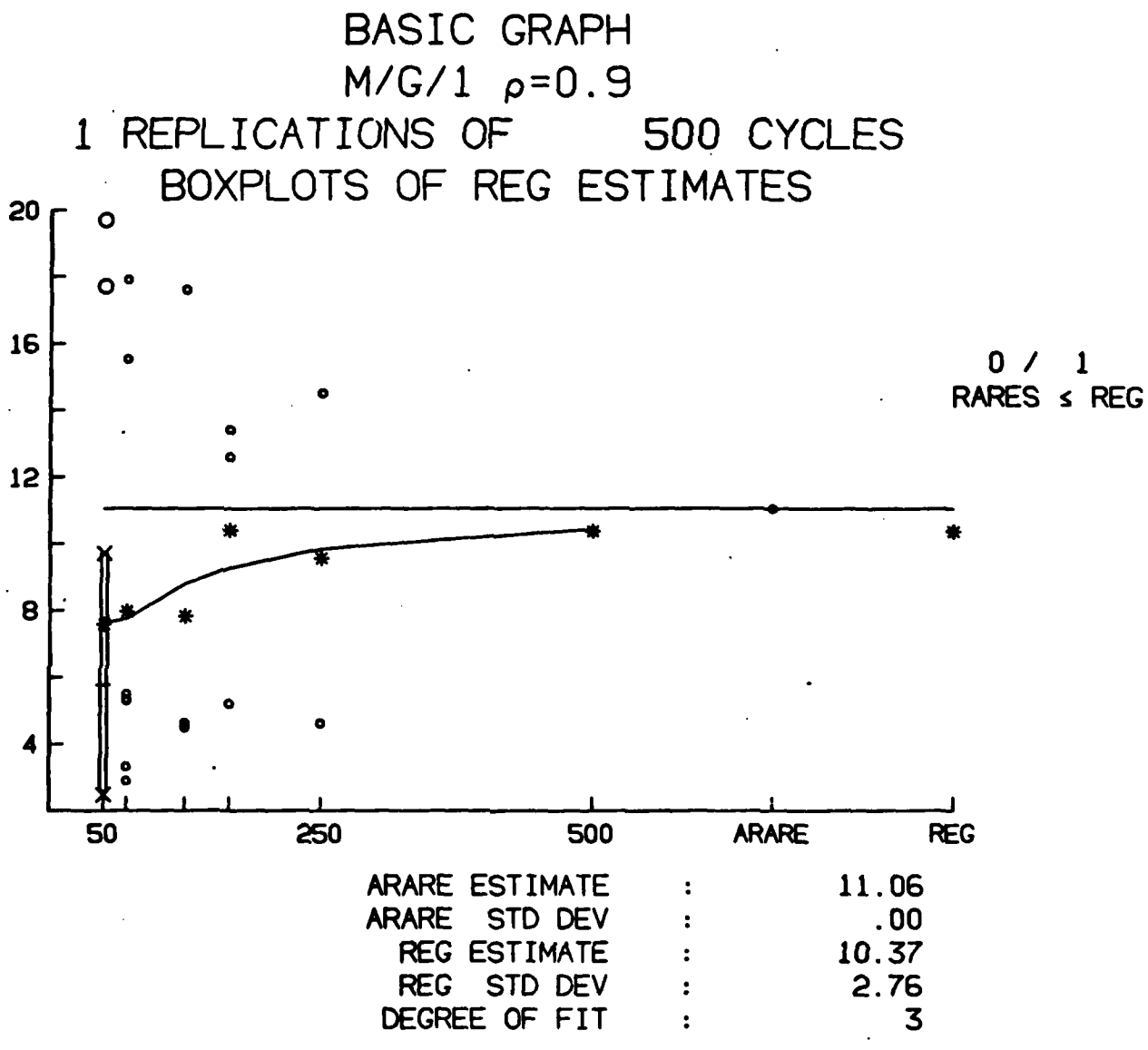


FIGURE 4A. Example 2. Stage 1, steps 1 and 2. N_0 set at 500 cycles. A decision has been made on the basis of similar graphs that d = 3 is probably needed.

Interpretation. Because of the skewed service time, there is more variability in this queue than in the M/M/1 queue. The regression curve is definitely still rising so that n = 500 may be too small. Also the precision is about 50%, so that about 25 times more cycles are needed. An exploratory Retrenched Graph is needed.

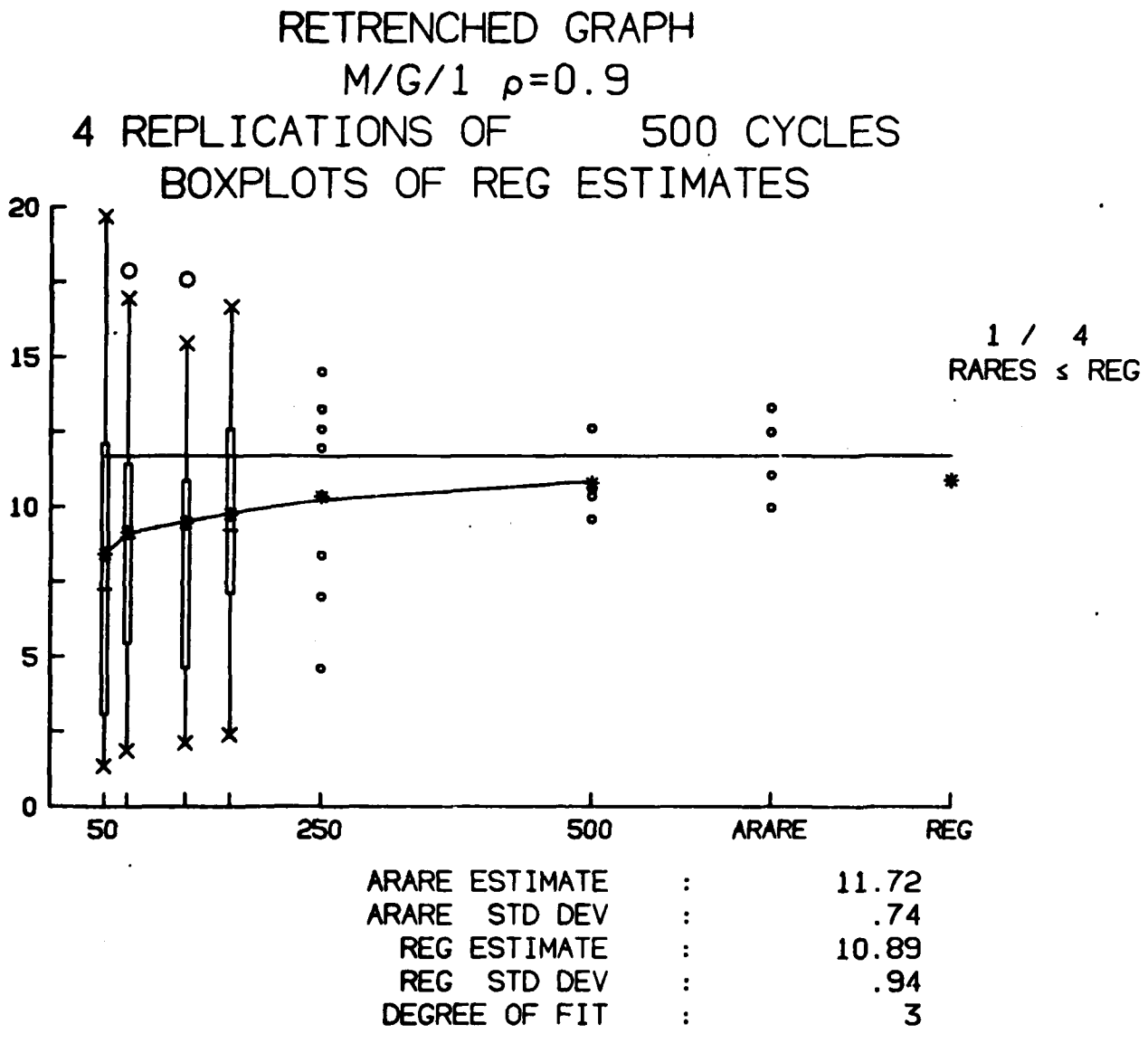


Figure 4B. Example 2. Stage II, step 3a. $N_0 = 2000$ and $n = 500$.

Interpretation. There is clearly still bias in the $\hat{re}(j,\overline{500})$'s and a need for more precision. Four times as many cycles are possibly needed; this is the "otherwise" of Stage II, Step 3c.

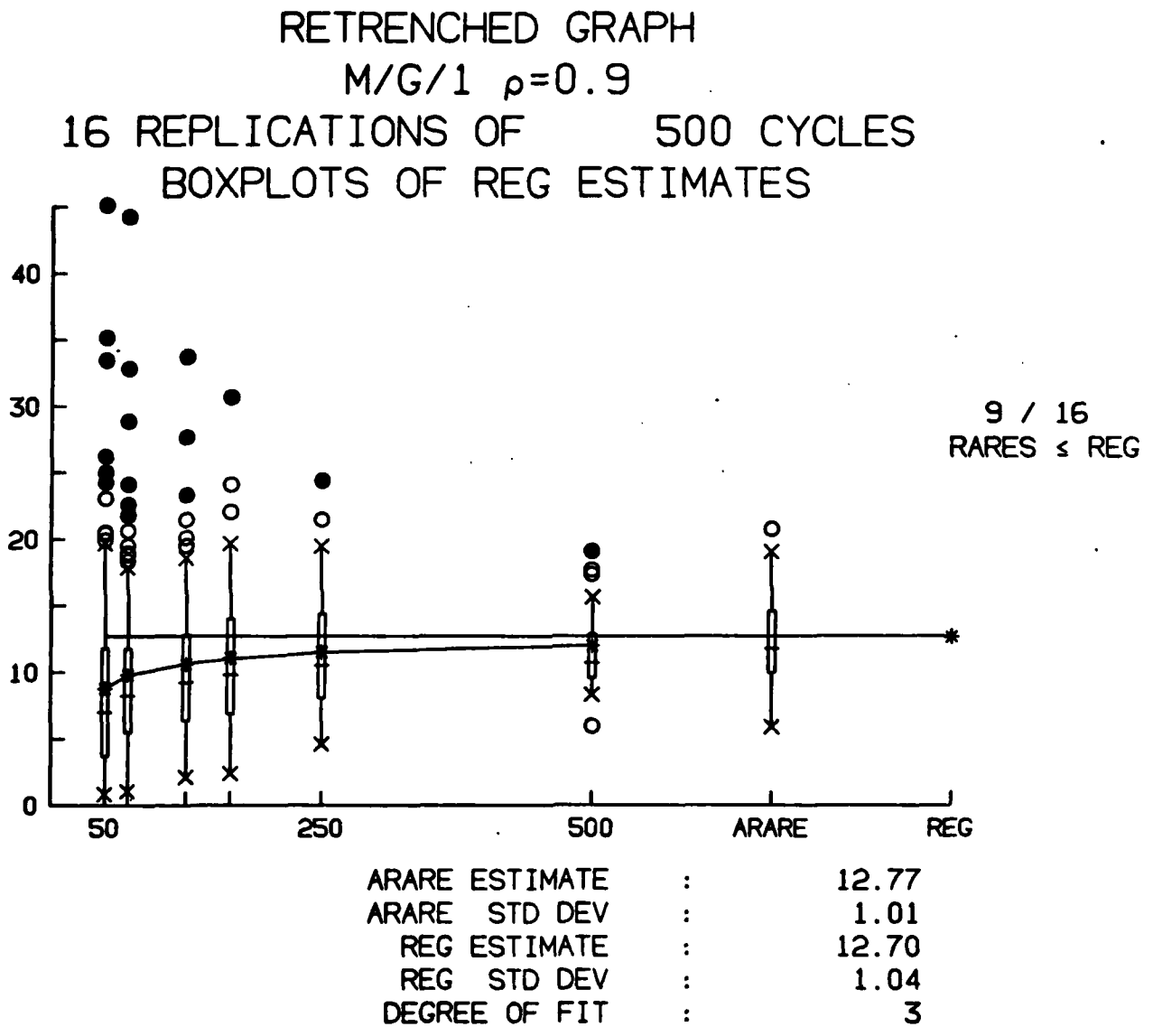


Figure 4C. Example 2. Stage II, Steps 1 through 3 with $N_0 = 8000$ and $n = 500$.

Interpretation. Note that the estimated standard deviations have increased from their values in Figure 4B. This is because several very long cycles were encountered; these show up in the upper outliers in the $\hat{re}(j,n_k)$ box plots. We are back at the "otherwise" of Stage II, Step 3c and increase N_0 to 16,000.

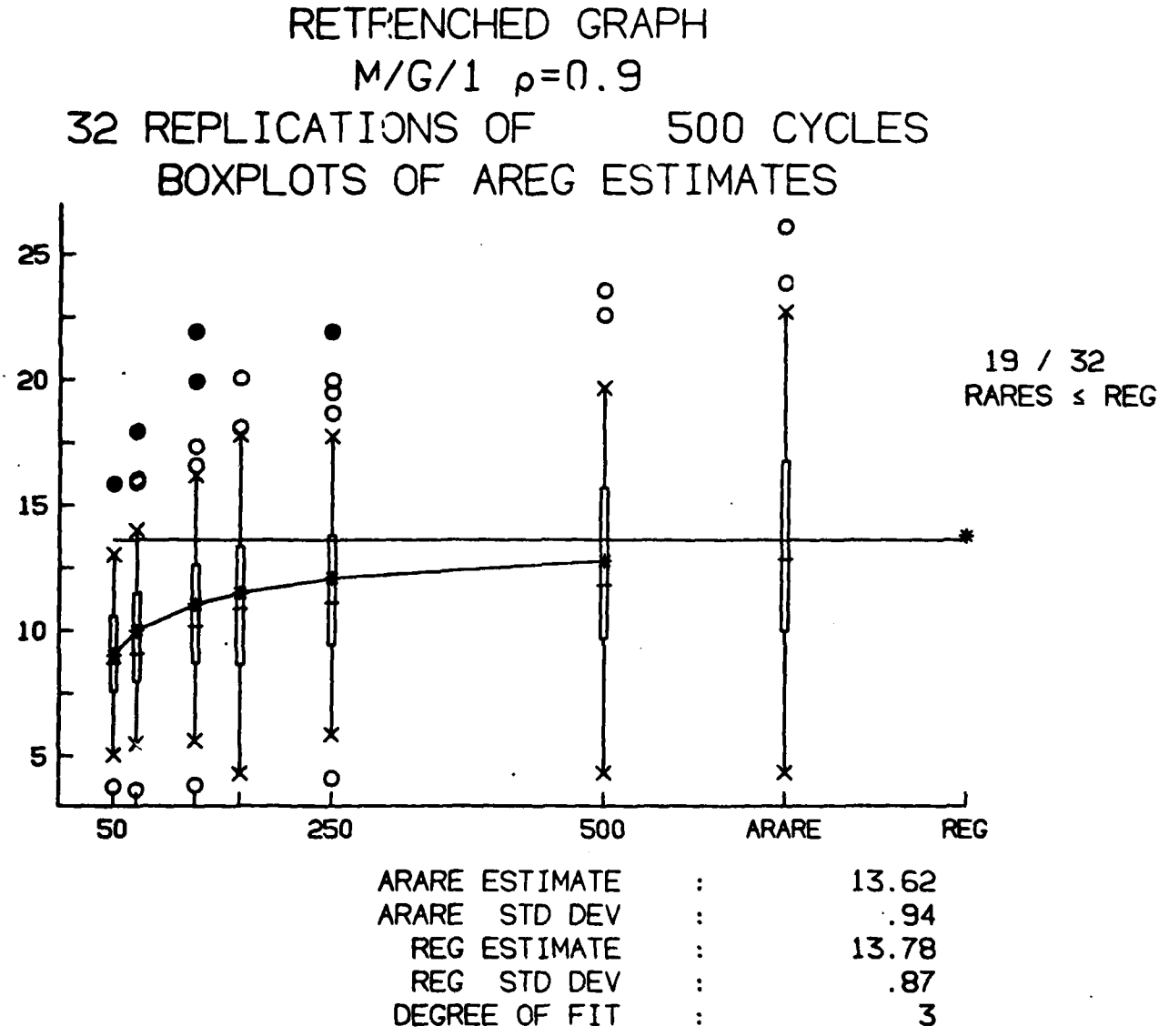


FIGURE 4D. Example 2, Stage II, Steps 1 through 3 with $N_0 = 16000$ and $n = 500$.

Interpretation. The point estimates have risen in value and the estimated standard deviations have dropped, but not quite enough to give 10% precision. Note that the true $E(W) = 13.42$. There is clearly still some bias in the $re(j,500)$'s; of the $rare(j,500)$'s, nineteen are less than or equal to $re(16000)$.

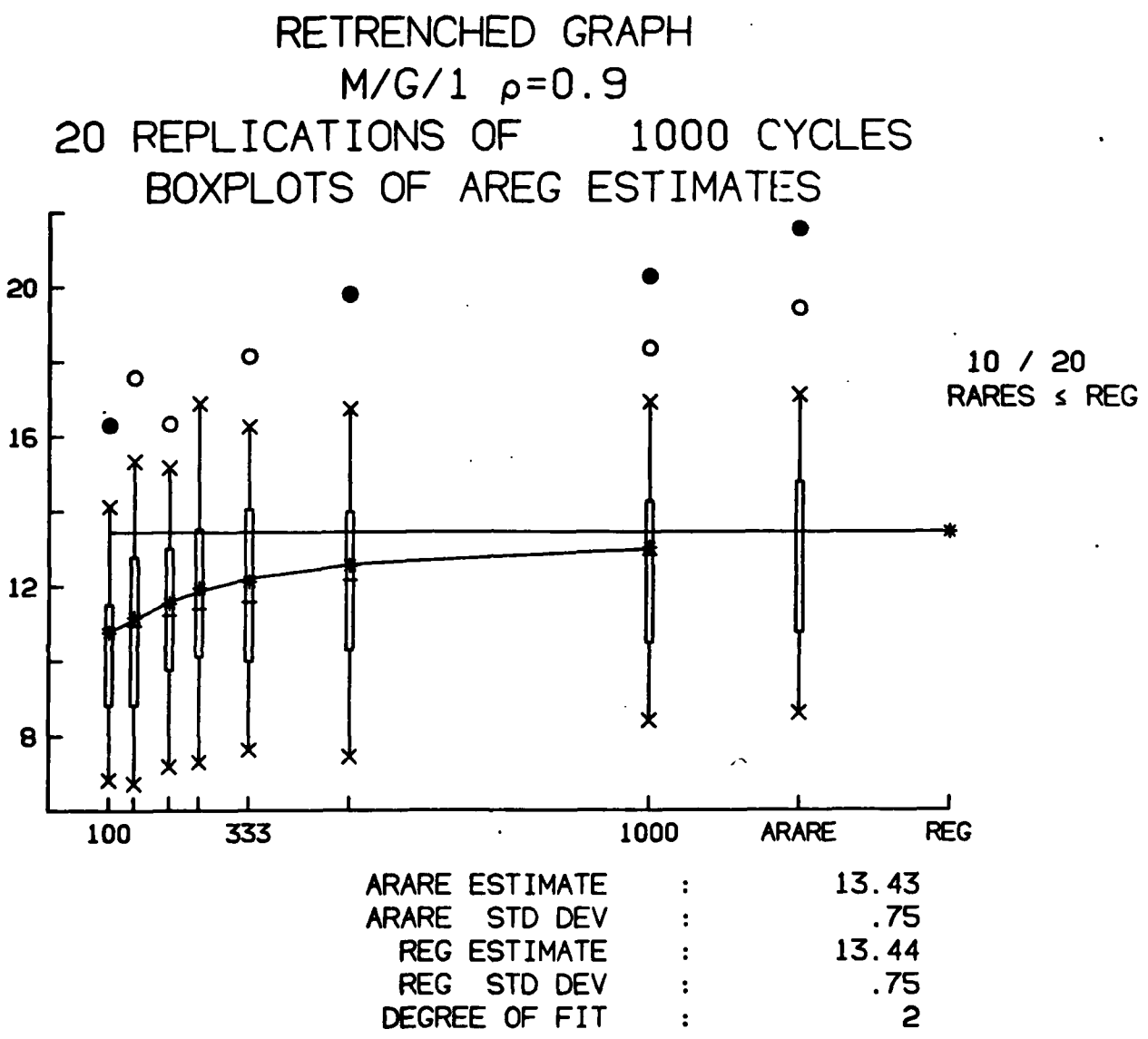
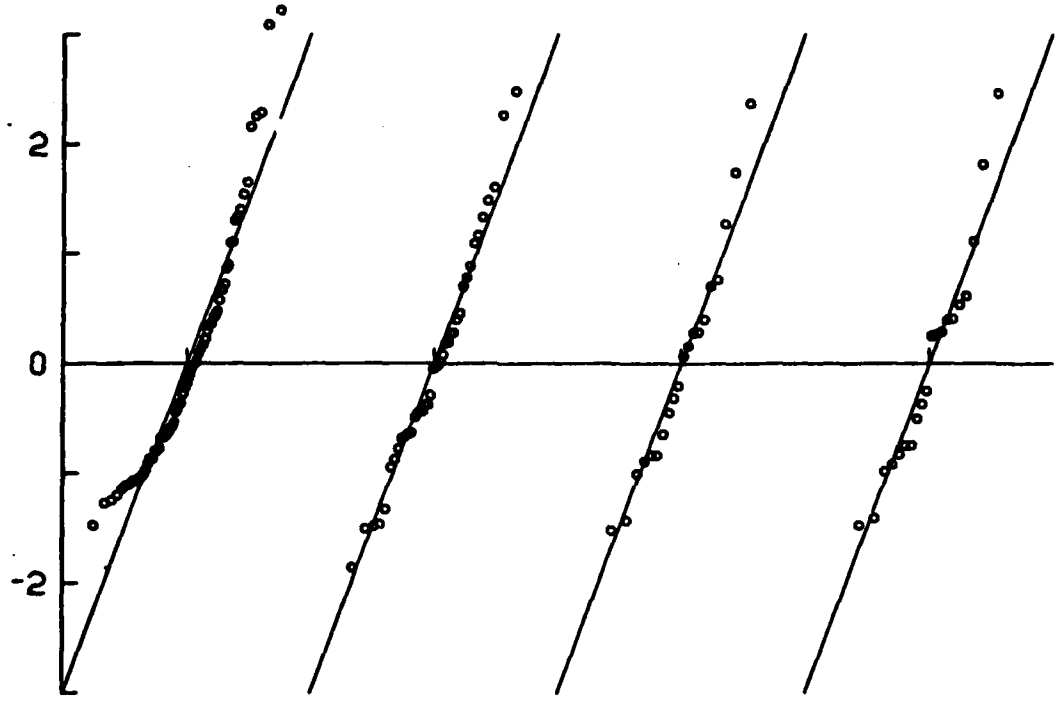


FIGURE 4E. Example 2. Stage II, steps 1 through 3 with N_0 increased to 20,000, the maximum number which computing constraints allowed. The degree of fit is d = 2, sufficient for the purpose, as compared to d = 3 in Figure 4D. The section length n is increased to n = 1000.

Interpretation. There is still indication of bias at n = 1000, but more particularly of outliers in the *re*(j,1000)'s and *rare*(j,1000)'s which need to be looked at in a running normal plot. Note that it is not quite possible to attain 10% accuracy.



	REG[250]	REG[500]	REG[1000]	RARE	
SHAPIRO-WILK	.890(.000)	.968(.409)	.959(.516)	.943(.265)	(SIG LEV
SKEW	1.142(.274)	.487(.387)	.606(.548)	.691(.548)	(SD SKE
KURTOSIS	1.154(.548)	-.097(.775)	-.171(1.095)	.056(1.095)	(SD KURT

FIGURE 4F. Example 2. Running normal plot accompanying Figure 4G.

Interpretation. The definite non-normality in the $re(j,250)$'s is no longer apparent in the values of the test statistics at n = 1000. However outliers are clearly apparent in the $re(j,1000)$'s and $rare(j,1000)$'s, suggesting, since m = 20, that n be doubled. (Stage II, step 4d.)

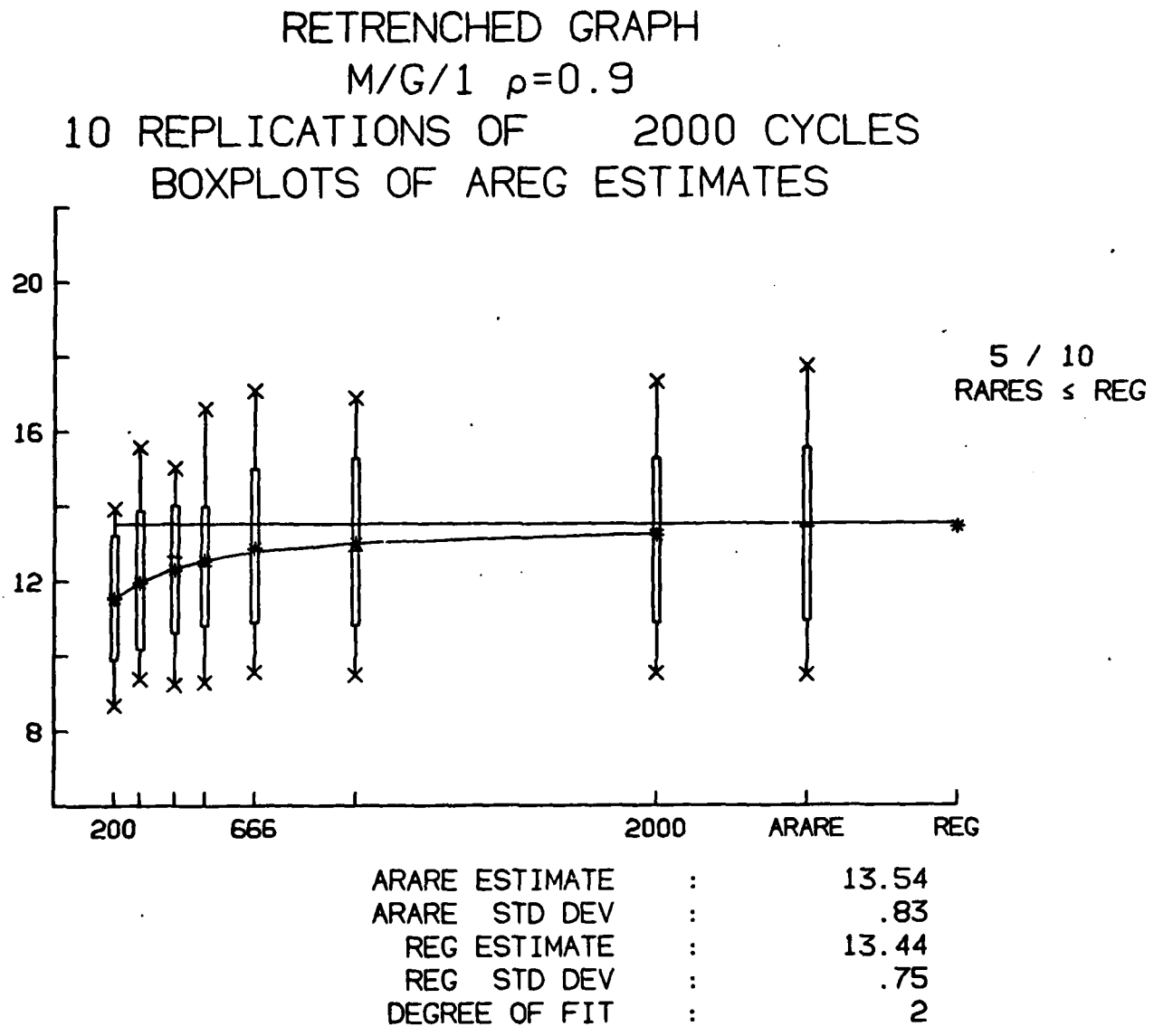
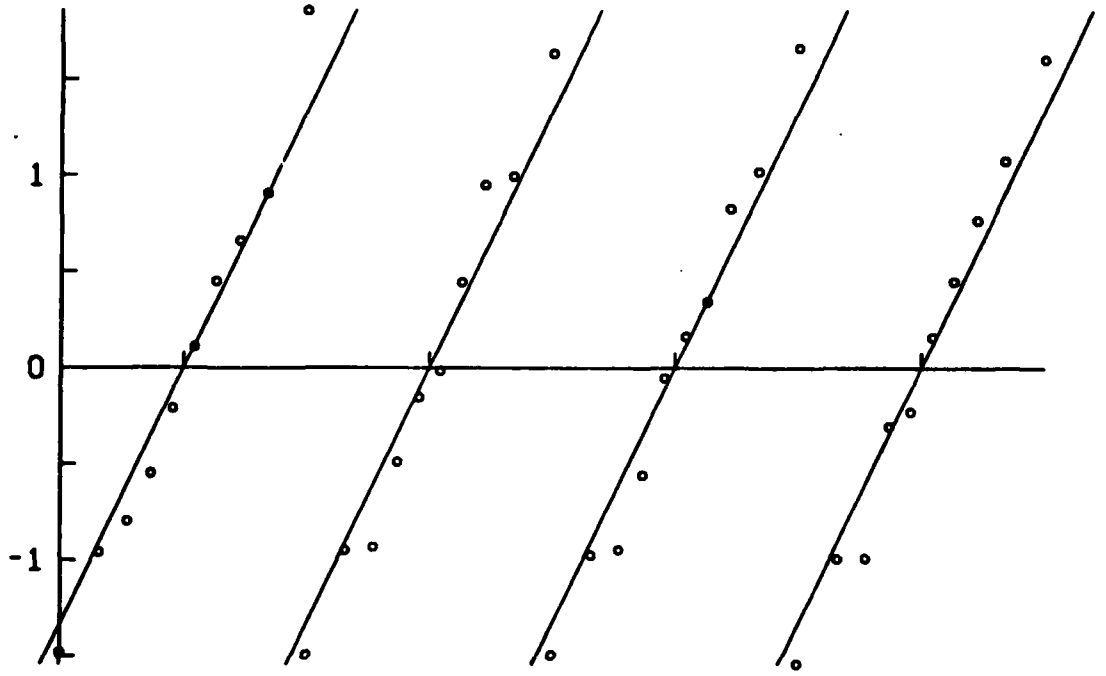


FIGURE 4G. Example 2. Stage II, step 4 with n increased to 2000, which keeps $m \geq 10$, as required. To achieve the desired precision 4000 more cycles would have been desirable.

Interpretation. The bias has been reduced but more importantly the heavy tails in the estimates have disappeared (as compared to Figure 4E) at all n_k's. To confirm these results a running normal plot is created.

M/G/1 ρ=0.9 NORMAL PLOTS
10 REPLICATIONS OF 2000 CYCLES



	AREG[500]	AREG[1000]	AREG[2000]	RARE	
SHAPIRO-WILK	.984(.982)	.971(.890)	.977(.944)	.979(.952)	(SIG LEV)
SKEW	.319(.775)	.143(.775)	.118(.775)	.032(.775)	(SD SKEW)
KURTOSIS	-.671(1.549)	-1.090(1.549)	-1.036(1.549)	-1.048(1.549)	(SD KURT)

FIGURE 4H. Example 2. Final check for normality with n = 2000 cycles. Note that the plots at 500 and 1000 are of the *arare*(j,n_k)'s.

Interpretation. The test statistics and the normal plots confirm the apparent normality of the estimates at n = 2000, although the sample size m = 10 is very small. The estimate *arare*(10, 2000) would be preferred to *are*(10,2000) because of bias considerations.

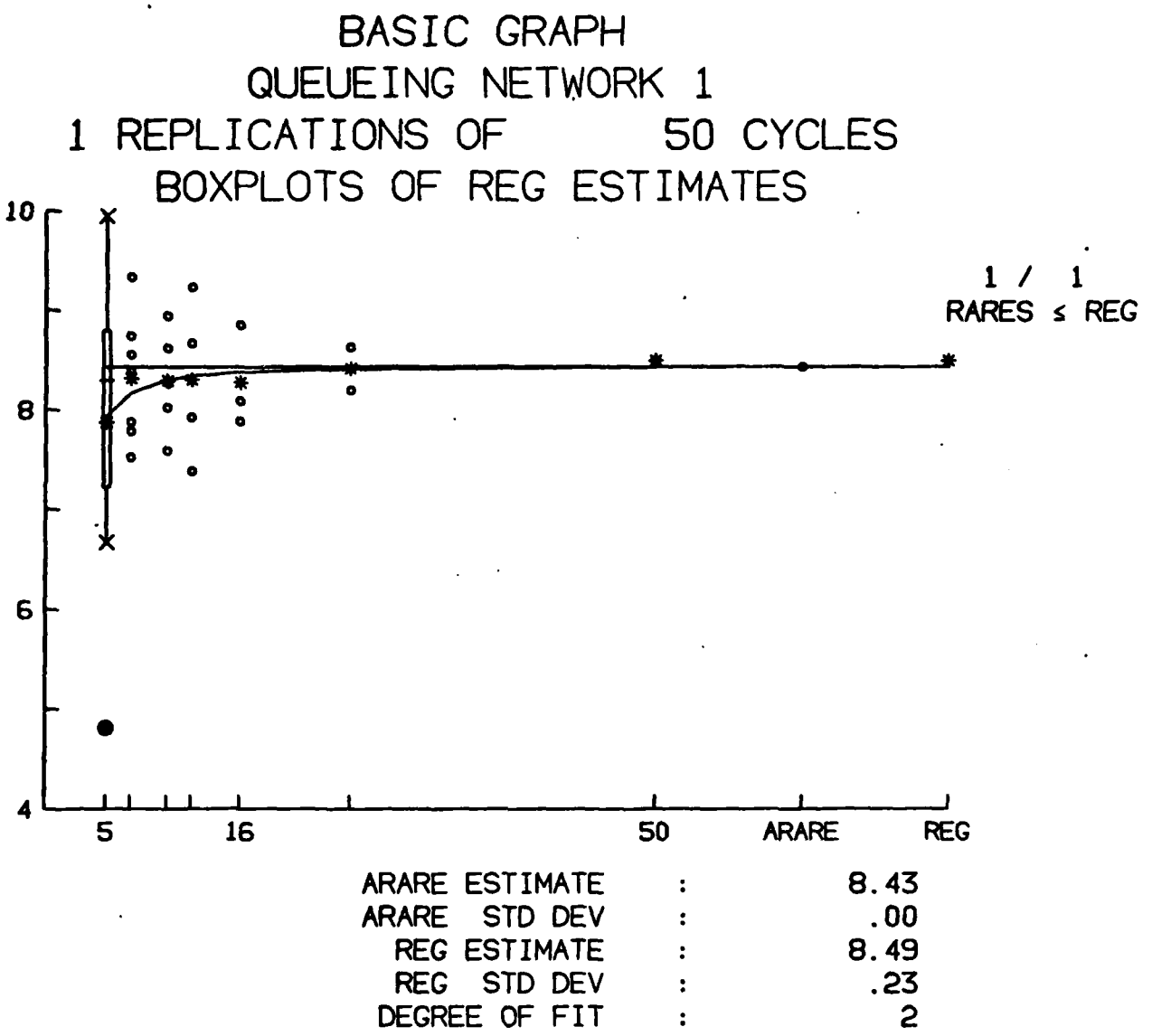


FIGURE 5A. Example 3. Stage I, steps 1, 2 and 3. N_0 set at 50 cycles: Degree of fit is d = 2.

Interpretation. This initial basic graph shows that the bias is probably very low at n = 50 and that the accuracy is already less than 6%. Clearly the $\hat{re}(j,5)$'s are non-normal but there is insufficient data to judge the distributions of the other estimates. Since cycles are relatively expensive, in practice a decision to terminate the situation at this point would be made. Thus the assumptions of low bias and normality are not confirmed. For purposes of exposition, however, the 50 cycles are replicated 25 times for the next figure.

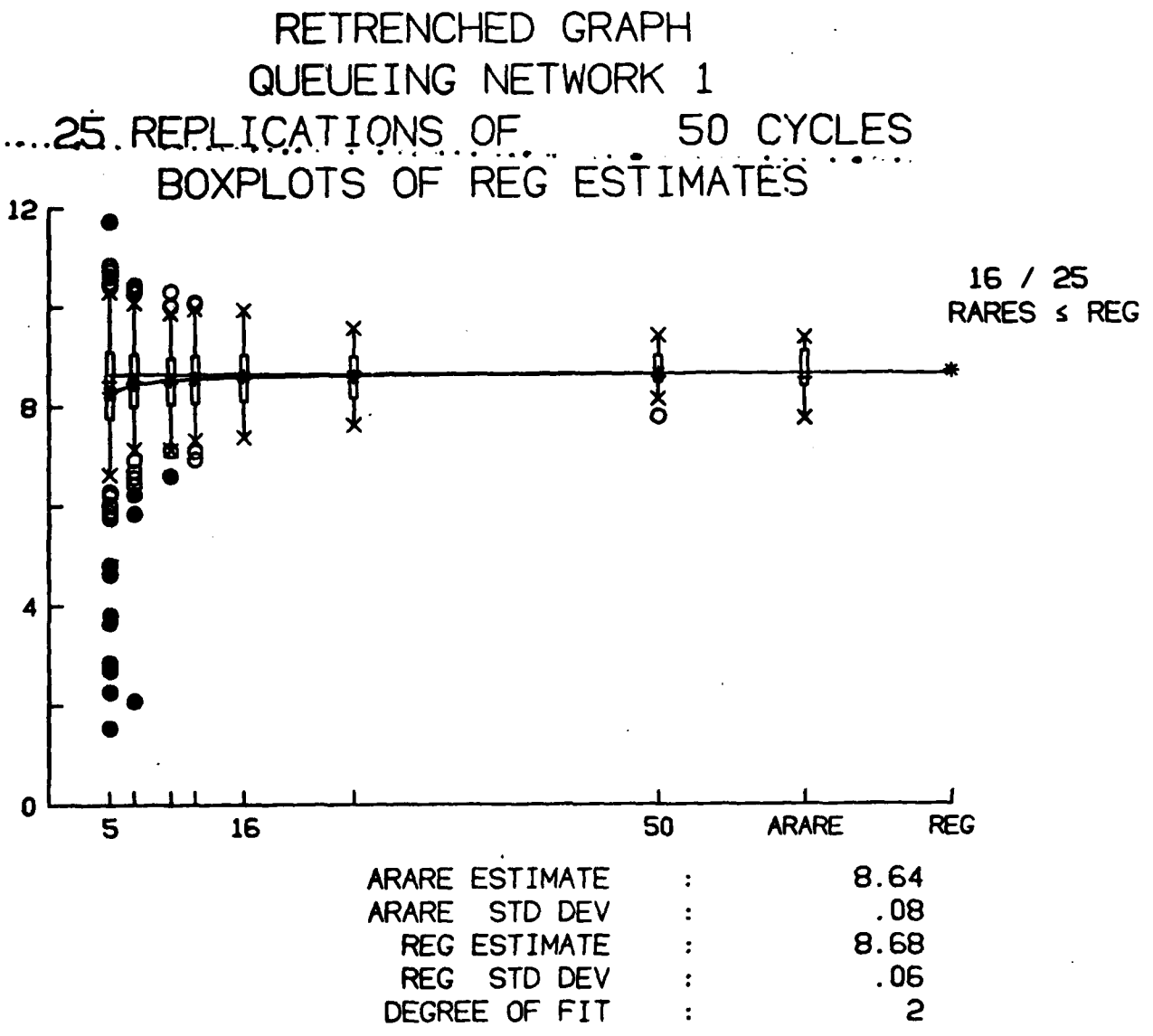


FIGURE 5B. Example 3. Stage II. N = 1250.

Interpretation. This retrenched graph shows clearly that bias is nonexistent at n = 50. Note that 16 of the *rare*(j,50)'s are less than *re*(1250). Note too the extreme negative skewness of the *re*(j,50)'s; this seems to have disappeared by n = 50 both for the *re*(j,50)'s and *rare*(j,50)'s. The very flat bias curve and the slight asymmetry in the distribution of the *rare*(j,50)'s suggest that the degree of fit in the regression could be cut back to one.

7. EMPIRICAL RESULTS

In this section we empirically compare the relative properties of the estimate $re(mn)$ and $arare(m,n)$ for the single server queues. These studies give the results alluded to in Section 3; namely that the $arare(m,n)$ estimate has low bias, and variance and mean squared error (mse) comparable to those of the regenerative estimate. Furthermore confidence intervals formed using $arare(m,n)$ and $srare(m,n)$ are valid, i.e. the confidence intervals contained the true value with the correct, prespecified probability. Thus even for large samples, a user looses nothing by switching from the regenerative to the *arare* estimate. The substantial benefit of doing so is that the user may easily judge what constitutes a "large sample" and at what point the asymptotic theory is valid. In addition $arare(m,n)$ and $srare(m,n)$ were less correlated with one another than were the regenerative point and standard deviation estimates, $re(mn)$ and $\hat{\sigma}(nm)$.

The above mentioned experiments, the results of which are compiled in Tables 1, 2, and 3, were conducted as follows. The systems tested, as in Examples 1 and 2, were highly congested M/M/1 and M/G/1 queues ($\rho = 0.9$). For each system and several combinations of m sections, n cycles per section and subsection lengths $\{n_k\}$, R independent simulations of $m \times n$ cycles were performed. These queues were simulated using the FORTRAN language and the random number generator described in Learmonth and Lewis (1973). Different seeds for the generator were used for different systems and different values of $m \times n$. On each replication the following quantities were computed: $re(nm)$, $arare(m,n)$, $\overline{are}(mn_k,n_k)$, $\hat{\sigma}(mn)$, $srare(m,n)$ and $\overline{sare}(mn_k,n_k)$. The entries in the tables are best explained by an example.

Consider the first row of Table 2. Let $re_i(4000)$ and $\hat{\sigma}_i(4000)$, for $i = 1,\ldots,R$, be respectively the realizations of the regenerative point and standard deviation estimates based on the 4000 cycles of the ith replication ($m = 8$, $n = 500$). The column labeled $\tilde{W}$ contains the sample average, say $\tilde{W}_{re}$, of $\{re_i(4000), i = 1,\ldots,R\}$; $\tilde{W}_{re} = \sum_{i=1}^{R} re_i(4000)/R = 8.76$. This is an unbiased estimate of

$E(re(4000))$, whose value is unknown. The sample standard deviation of $\tilde{W}_{re}$ is $s(\tilde{W}_{re}) = [\sum_{i=1}^{R}\{re_i(4000) - \tilde{W}_{re}\}^2/R(R-1)]^{1/2} = 0.09$ and is listed in the column labeled $s(\tilde{W})$. Thus a 90% confidence interval for $E(re(4000))$ is $8.76 \pm 1.645 \times 0.09 = 8.76 \pm 0.15$. Since for this example $E(W) = 9.00$ does not fall within this confidence interval we conclude that $re(4000)$ is significantly biased. An estimate of the mean squared error of $re(4000)$, $MSE_{re} = \sum_{i=1}^{R} (re_i(4000) - E(W))^2/R = 0.87$, is listed in the MSE column of Table 2. Similarly estimates of the correlation and Spearman, or rank, correlation coefficients (Snedecor and Cochran (1967)) between $re(4000)$ and $\hat{\sigma}(4000)$, 0.77 and 0.67, respectively, are listed.

Using $re_i(4000)$ and $\hat{\sigma}_i(4000)$ alleged 90% confidence intervals, $re_i(4000) \pm 1.645\hat{\sigma}_i(4000)$, for $E(W)$ were also formed. The fraction of these confidence intervals that actually contained $E(W)$ is reported in the column labeled "90% coverage." This fraction, called "(90%) coverage," should be close to 0.90 if in fact valid confidence intervals are being formed. Any coverage less than or equal to 0.85 in Table 2 (with $R = 100$) and any coverage less than or equal to 0.865 in Tables 1 and 3 (with $R = 200$) are significantly less than 0.90 at the 0.10 level.

Table 1 focuses on the effects of changing the set of subsection lengths $\{n_k\}$ on the *arare* estimate. The table reports results of simulations of 2000 cycles of the M/M/1 queue with $\rho = 0.9$. The graphs of Section 6 show this to be a relatively short run length. The estimates in this table used a polynomial of degree 3 in the regression; degrees 1 and 2 are omitted since they showed the same trends but with higher bias and lower coverage. Notice that the *arare* estimates for which the smallest value of n_k is $n/10$ (subsection sets S_3, S_4, S_7 and S_8 in the table) show less bias and higher coverage than those with values of $n_k < n/10$ (sets S_1, S_2, S_5 and S_6). The values of n_k in subsection sets S_3 and S_7 are $n/10$, $[n/7]$, $n/5$, $n/4$, $[n/3]$, $n/2$, and n, which is the recommended set of points. Due to the high correlation between the

estimates for different values of n_k, there is little, if any, advantage in using more values of n_k than are in the recommended sets (compare sets S_3 versus S_4 and S_7 versus S_8). The *arare* estimate is relatively insensitive to the subsection lengths as long as they are not too small. However, in some cases adding points that are not divisors of n will increase the *arare*'s variance. The set of points recommended above is relatively small and has consistently yielded satisfactory results.

Tables 2 and 3 report the results of regenerative simulations with more cycles than the simulation of Table 1. The main trends to notice in these tables are that:

(i) the bias, standard deviation, coverage and mse of *arare*(m,n) is comparable to that of *re*(mn);

(ii) *arare*(m,n) shows lower bias, lower mse and higher (truer) coverage than *are*(m,n),

(iii) for fixed values of $m \times n$ it is preferable to have small values of m, the number of sections, in terms of bias, mse, coverage and correlation between point and standard deviation estimates, and

(iv) the correlation between *re*(mn) and $\hat{\sigma}$(mn) is higher than that between *arare*(m,n) and *srare*(m,n).

This last property warrants further discussion. As n_k gets large the *are*(j,m_k,n_k)'s converge to normal random variables and since *rare*(j,n) is a linear combination of $\{$*are*(j,m_k,n_k), $k = 1,\ldots,K\}$, it too converges to normality. Thus *arare*(m,n) and *srare*(m,n) are asymptotically independent. The nonzero correlations in Tables 2 and 3 indicate that for these highly congested queues the convergence to normality is very slow and that the independence of mean and variance is quite sensitive to the normality assumption. Of course these correlations are seen in Tables 2 and 3 to be less than the correlation between *re*(mn) and $\hat{\sigma}$(mn). Also non-parametric methods can be used with the set of m *rare*(j,n)'s.

The high correlation between $re(N)$ and $\hat{\sigma}(N)$, where $N = mn$, is of particular concern since sequential rules for determining simulation run lengths typically rely on the relative width of a confidence interval, a multiple of $\hat{\sigma}(N)/re(N)$, as a stopping criterion [Lavenberg and Sauer (1977)]. If $re(N)$ is smaller than usual, then because of the high correlation $\hat{\sigma}(N)$ will also be smaller than usual. This leads to a "low miss" of the true answer, i.e. an unusually small confidence interval centered about an unusually small point estimate. Since sequential rules are specifically designed to stop on small confidence intervals, the rule inevitably stops a large proportion of low misses. This will not be a problem with the graphical methodology given here, since in addition to the lower correlation between $arare(m,n)$ and $srare(m,n)$, the distributional information displayed will guard against the skewed data which causes the low miss.

8. SUMMARY AND FURTHER WORK

The statistical and graphical methodology given in this paper has been shown to be very effective in verifying the bias and normality properties of regenerative estimates and in sectioning up a regenerative simulation so that more reliable standard deviation estimates and confidence interval estimates can be obtained than with the usual regenerative methodology.

Although the design of the regression for the regression-adjusted regenerative estimates has been shown to be robust in the cases considered, analytic results to confirm this in a broader class of simulation situations will be pursued. The regression-based variance estimate for the regenerative estimate also needs to be explored further. This will be useful when applications of the methodology to other statistical and simulation output situations is considered.

Another point to be explored is the use of robust regression techniques, as at 3ii. This could be fruitful because the $re(j,n_k)$'s are highly non-normal, as can be seen in Figure 4C at $n_k = 50$.

Finally we note that the regression-adjusted estimation procedure and its associated graphics can be applied in many other areas of statistics, and in particular in the statistical analysis of simulation outputs. Thus if one has an estimator $\hat{\theta}(n) = \hat{\theta}(X_1,\ldots,X_n)$ of a parameter θ from an i.i.d. sample $X_1,\ldots,X_n$, similar estimates $\hat{\theta}(j,n_k)$, $j = 1,\ldots,m_k$, can be formed from m_k subsections of the sample for $k = 1,\ldots,K$ to give a regression-adjusted estimator, with graphics, just as it was done for the regenerative estimator $\mathcal{re}(n)$. This application, and multivariate extensions, will be explored elsewhere.

Table 1

Effects of Subsection Lengths on *Arare* Estimates of E(W) = 9.00 in the M/M/1 Queue with $\rho = 0.9$. Here R = 200.

Estimate	Set of Subsection Lengths for *Arare*	$\bar{W}$	$S(\bar{W})$	90% Coverage	MSE
re(2000)		8.725	.086	.810	1.552
are(10,200)		8.086	.071	.715	1.847
are(20,100)		7.513	.061	.460	2.948
arare(10,200)	S_1	8.466	.085	.820	1.707
arare(10,200)	S_2	8.459	.075	.820	1.424
arare(10,200)	S_3	8.753	.090	.865	1.661
arare(10,200)	S_4	8.673	.088	.850	1.659
arare(20,100)	S_5	8.172	.077	.725	1.864
arare(20,100)	S_6	8.129	.068	.730	1.691
arare(20,100)	S_7	8.466	.078	.790	1.485
arare(20,100)	S_8	8.459	.081	.780	1.592

Degree of polynomial fitted in regression was 3.

$S_1 = \{n_k : n_k = 5k,\ k = 1,\ldots,40\}$

$S_2 = \{n_k : n_k = 5, 10, 20, 28, 40, 50, 66, 100, 200\}$

$S_3 = \{n_k : n_k = 20, 28, 40, 50, 66, 100, 200\}$

$S_4 = \{n_k : n_k = 20 + (k-1)5,\ k = 1,\ldots, 37\}$

$S_5 = \{n_k : n_k = 4 + (k-1)3,\ k = 1,\ldots, 33\}$

$S_6 = \{n_k : n_k = 4, 7, 10, 14, 20, 25, 33, 50, 100\}$

$S_7 = \{n_k : n_k = 10, 14, 20, 25, 33, 50, 100\}$

$S_8 = \{n_k : n_k = 10 + (k-1)3,\ k = 1,\ldots, 31\}$

Table 2

Simulation Results for Estimates of E(W) = 9.00
in M/M/1 Queue with $\rho = 0.9$. Here R = 100.

Estimate	$\tilde{W}$	$S(\tilde{W})$	90% Coverage	MSE	Correlation $(\hat{W},\hat{\sigma}(W))$	Spearman Correlation $(\hat{W},\hat{\sigma}(W))$
re(4000)	8.76	.09	.84	.87	.77	.67
arare(8,500)(d=2)*	8.74	.09	.87	.87	.64	.48
arare(8,500)(d=3)	8.75	.09	.87	.88	.62	.49
are(8,500)	8.47	.08	.81	.95	na**	na
arare(16,250(d=2)	8.63	.09	.84	.87	.70	.58
arare(16,250)(d=3)	8.74	.10	.86	.98	.75	.65
are(16,250)	8.20	.08	.72	1.23	na	na
re(8000)	8.99	.07	.86	.50	.71	.72
arare(8,1000)(d=2)	9.00	.07	.93	.52	.48	.36
arare(8,1000)(d=3)	9.02	.07	.93	.53	.49	.34
are(8,1000)	8.83	.07	.88	.49	na	na
arare(16,500)(d=2)	8.94	.07	.88	.49	.56	.47
arare(16,500)(d=3)	8.99	.07	.87	.52	.62	.58
are(16,500)	8.66	.06	.75	.53	na	na
re(16000)	9.04	.05	.92	.22	.64	.59
arare(8,2000)(d=2)	9.05	.05	.93	.22	.50	.43
arare(8,2000)(d=3)	9.05	.05	.89	.22	.53	.48
are(8,200)	8.96	.05	.88	.20	na	na
arare(16,1000)(d=2)	9.04	.05	.93	.22	.47	.42
arare(16,1000)(d=3)	9.06	.05	.89	.24	.53	.48
are(16,1000)	8.88	.04	.87	.21	na	na
arare(32,500)(d=2)	9.00	.05	.91	.20	.54	.49
arare(32,500)(d=3)	9.04	.05	.91	.23	.59	.56
are(32,500)	8.71	.04	.79	.26	na	na

*d = degree of polynomial fitted in regression

**na:not available

subsection lengths $\{n_k\}$ = n/10, [n/7], n/5,n/4, [n/3], n/2, n

Table 3

Simulation Results for Estimates of $E(W) = 12.77$ in M/G/1 Queue with $\rho = 0.9$ ($\lambda = 1.0$, $\mu_1 = 0.5$, $\mu_2 = 2.0$, $p = 0.33$). Here $R = 200$.

Estimate	$\tilde{W}$	$S(\tilde{W})$	90% Coverage	MSE	Correlation $(\hat{W},\hat{S}(W))$	Spearman Correlation $(\hat{W},\hat{S}(W))$
re(5000)	12.62	.09	.810	1.73	.71	.77
arare(10,500)(d=2)*	12.56	.09	.865	1.78	.61	.60
arare(10,500)(d=3)	12.61	.10	.875	1.89	.62	.60
are(10,500)	12.03	.08	.770	2.02	na**	na
re(10000)	12.64	.07	.850	1.03	.70	.71
arare(10,1000)(d=2)	12.66	.07	.880	1.00	.50	.45
arare(10,1000)(d=3)	12.72	.07	.910	1.01	.52	.46
are(10,1000)	12.39	.07	.860	1.04	na	na
arare(20,500)(d=2)	12.51	.07	.825	1.16	.60	.58
arare(20,500)(d=3)	12.60	.08	.825	1.22	.65	.63
are(20,500)	12.00	.07	.700	1.54	na	na
re(20000)	12.78	.05	.855	.51	.72	.73
arare(5,4000)(d=2)	12.80	.05	.885	.51	.39	.36
arare(5,4000)(d=3)	12.80	.05	.880	.53	.37	.35
are(5,4000)	12.71	.05	.870	.50	na	na
arare(10,2000)(d=2)	12.78	.05	.880	.50	.49	.49
arare(10,2000)(d=3)	12.81	.05	.880	.53	.52	.52
are(10,2000)	12.63	.05	.820	.50	na	na
arare(20,1000)(d=2)	12.73	.05	.860	.52	.52	.49
arare(20,1000)(d=3)	12.79	.05	.855	.56	.55	.52
are(20,1000)	12.47	.05	.815	.59	na	na

*d = dgree of polynomial fitted in regression

**na: not available

for $n = 500$, $\{n_k\} = 50, 71, 100, 125, 166, 250, 500$

$n = 1000$, $\{n_k\} = 50, 100, 125, 166, 200, 250, 333, 500, 1000$

$n = 2000$ $\{n_k\} = 100, 166, 200, 250, 400, 500, 666, 1000, 2000$

$n = 4000$ $\{n_k\} = 400, 571, 800, 1000, 1333, 2000, 4000$

Table 4

Summary of Notation

$re(p)$ --regenerative estimate based on p cycles; $= \sum_{k=1}^{p} Y_k / \sum_{k=1}^{p} \tau_k$.

$re(\ell,p)$ --realization of $re(p)$ on the ℓth subsection out of the m_k subsections in a replication.

$are(m_k,n_k)$ --average regenerative estimate from subsections of length m_k; $= \sum_{j=1}^{m_k} re(j,n_k)/m_k$.

$are(j,m_k,n_k)$ --realization of $are(m_k,n_k)$ on the jth replication.

$\overline{are}(mm_k,n_k)$ --overall average regenerative estimate for mm_k subsections of length n_k; $= \sum_{j=1}^{m} are(j,m_k,n_k)/m$.

$s\overline{are}(mm_k,n_k)$ --sample variance of $\overline{are}(mm_k,n_k)$; $= \{ \sum_{j=1}^{m} (are(j,m_k,n_k) - \overline{are}(mm_k,n_k))^2/m(m-1)\}^{1/2}$.

$rare(n)$ --regression adjusted regenerative estimate using n cycles.

$rare(j,n)$ --realization of $rare(n)$ on the jth replication of a section of length n.

$arare(m,n)$ --average regression adjusted regenerative estimate from m sections of length n; $= \sum_{j=1}^{m} rare(j,n)/m$.

$srare(m,n)$ --sample standard deviation of $arare(m,n)$; $= \{ \sum_{j=1}^{m} (rare(j,n) - arare(m,n))^2/m(m-1)\}^{1/2}$.

REFERENCES

Baskett, F., Chandy, K. M., Muntz, R. R. and Palacious, F. G. (1975). Open, closed, and mixed networks of queues with different classes of jobs. J. Assoc. Comput. Mach. 22, 248-270.

Crane, M. A. and Iglehart, D. L. (1975). Simulating Stable Stochastic Systems, III: Regenerative processes and discrete-event simulations. Operat. Res. 23, 33-45.

Crane, M. A. and Lemoine, A. J. (1977). An Introduction to the Regenerative Method for Simulation Analysis. Springer-Verlag, New York.

Efron, B. and Stein, C. (1978). The jackknife estimate of variance. Stanford Statistics Dept. Technical Report No. 120.

Fishman, G. S. (1977). Achieving specific accuracy in simulation output analysis. Comm. ACM 20, 310-315.

Goodman, A. S., Lewis, P.A.W. and Robbins, H. E. (1971). Simultaneous estimation of large numbers of extreme quantities in simulation experiments. Naval Postgraduate School Report NPS55Lw71122A.

Heidelberger, P. (1978). Steady State estimation from finite length simulations. IBM Research Report RC7451, Yorktown Heights, New York.

Iglehart, D. L. (1978). The regenerative method for simulation analysis. Current Trends in Programming Methodology, Vol. III: Software Engineering. K. M. Chandy and R. T. Yeh (eds.). Prentice-Hall, Inc. Englewood Cliffs, New Jersey.

Iglehart, D. L. (1975). Simulating stable stochastic systems, V: Comparison of ratio estimators. Naval Res. Logist. Quart. 22, 553-565.

Kleinrock, L. (1975). Queueing Systems, Volume 1: Theory, John Wiley and Sons, Inc., New York.

Lavenberg, S. S. and Sauer, C. H. (1977). Sequential stopping rules for the regenerative method of simulation. IBM J. Res. Develop. 21, 545-558.

Learmonth, G. P. and Lewis, P.A.W. (1973). Naval Postgraduate School random number generator LLRANDOM. Naval Postgraduate School Report NPS55Lw73061A, Monterey, CA

Lilliefors, H. W. (1967). On the K-S test for normality with mean and variance unknown. J. Amer. Stat. Assoc. 62, 399-402.

McGill, R., Tukey, J. W. and Larsen, W. A. (1978). Variations of Box Plots. Amer. Statistician, 32, 1, 12-16.

McNeil, D. R. (1977). Interactive Data Analysis. John Wiley and Sons, Inc., New York.

Meketon, M. S. (1979). Analysis of the regenerative method for finite length simulation Technical Report 400, Department of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York.

Miller, R. G., Jr. (1974). The jackknife--a review. Biometrika 61, 1-15.

Mosteller, F. and Tukey, J. W. (1977). Data analysis and regression. Reading-Mass. Addison-Wesley.

Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). Biometrika, 52, 591-611.

Shapiro, S. S. and Wilk, M. B. (1968). Approximations for the null distribution of W statistics. Technometrics 10, 861-866.

Shapiro, S. W., Wilk, M. and Chen, H.J. (1968). A comparative study of various tests for normality. J. Amer. Stat. Assoc. 63, 1343-73.

Snedecor, G. W. and Cochran, W. G. (1967). Statistical Methods, Sixth Edition: The Iowa State University Press, Ames, Iowa.

Tukey, J. W. (1977). Exploratory Data Analysis. Addison-Wesley Publishing Co., Inc. Reading, Mass.

Wilk, M. B. and Gnandesikan, R. (1968). Probability plotting methods for the analysis of data. Biometrika, 55, 1-17.

INITIAL DISTRIBUTION LIST

	NO. OF COPIES
Defense Technical Information Center Cameron Station Alexandria, VA 22314	12
Library, Code 0142 Naval Postgraduate School Monterey, CA. 93940	2
Library, Code 55 Naval Postgraduate School Monterey, CA 93940	1
Dean of Research, Code 012A Naval Postgraduate School Monterey, CA 93940	1
Statistics and Probability Program Code 436, Attn: E.J. Wegman Office of Naval Research Arlington, VA 22217	3
Dr. P. Heidelberger IBM, P. O. Box 218 T. J. Watson Res. Ctr Yorktown Heights, N.Y. 10598	10
Naval Postgraduate School Monterey, CA 93940	
Attn: P.A.W. Lewis, Code 55Lw	255
R. J. Stampfel, Code 55	1


DAT
FILM